numero 121
settembre 1978
anno XI
volume XXI

LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

Progetti alternativi per il motore d'automobile

Più che i vantaggi, sono evidenti i difetti dei motori progettati, in alternativa al tradizionale motore a ciclo Otto, per realizzare scarichi meno inquinanti e un migliore rendimento termodinamico

di David Gordon Wilson

Il convenzionale motore d'automobile con accensione a scintilla possiede pregi intrinseci, che risultano ancora più evidenti quando si vogliono costruire motori alternativi. Questi pregi comprendono un rendimento notevole (specialmente a carico parziale), una messa in moto facile, una quantità accettabile di gas di scarico (con regolatori) e un'esigenza minima di materiali di costruzione costosi (con conseguente basso costo di produzione). I sistemi di propulsione che il più delle volte vengono indicati come potenzialmente più promettenti del motore con accensione a scintilla (e cioè a ciclo Otto) per quel che riguarda il consumo di carburante e l'emissione di gas di scarico, sono il motore a ciclo Diesel, il motore a vapore (Rankine), la turbina a gas (motore di Brayton), il motore Stirling e la trazione elettrica a batteria. Potenzialmente tutti questi motori sono in grado di migliorare in misura notevole rispetto al loro stato di sviluppo attuale, e questo loro potenziale li renderebbe rivali estremamente allettanti per una futura produzione di serie, anche se la trazione elettrica a batteria potrebbe avere caratteristiche d'impiego solo molto limitate. Forse i problemi principali che dovrà affrontare chi prende le decisioni in sede governativa, nel caso in cui si voglia incoraggiare lo sviluppo di un motore migliorato, sono il fervore missionario che anima i sostenitori dei vari motori e i criteri eccessivamente semplicistici con cui si tende a giudicare i motori in questione. La scelta di un motore nuovo è un fatto estremamente complesso. Da parte mia, dopo aver esaminato con grande attenzione i pro e i contro dei vari motori in concorrenza, suggerirò il tipo di politica governativa che, a mio parere, ha maggiori probabilità di identificare, in modo rapido ed economico, il o i motori (ammesso che ve ne siano) in grado di soppiantare l'invenzione, vecchia ormai di 102 anni, di Nikolaus August Otto.

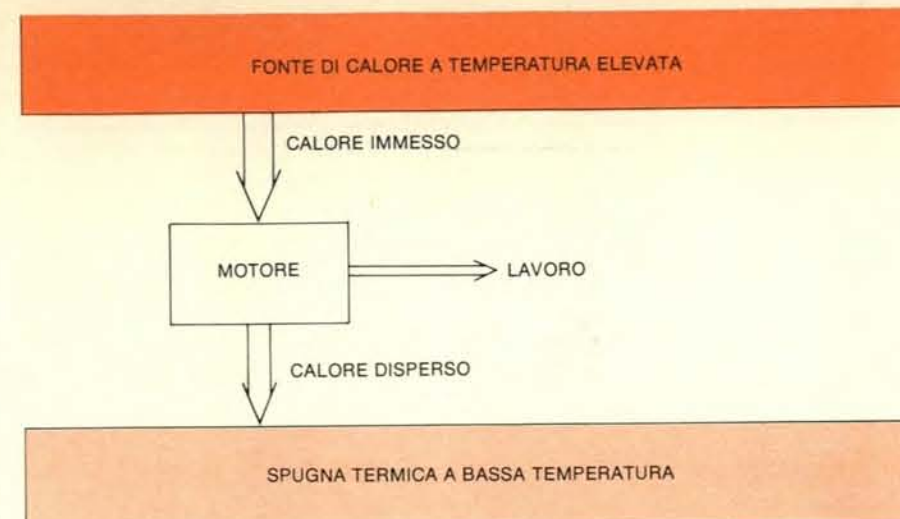
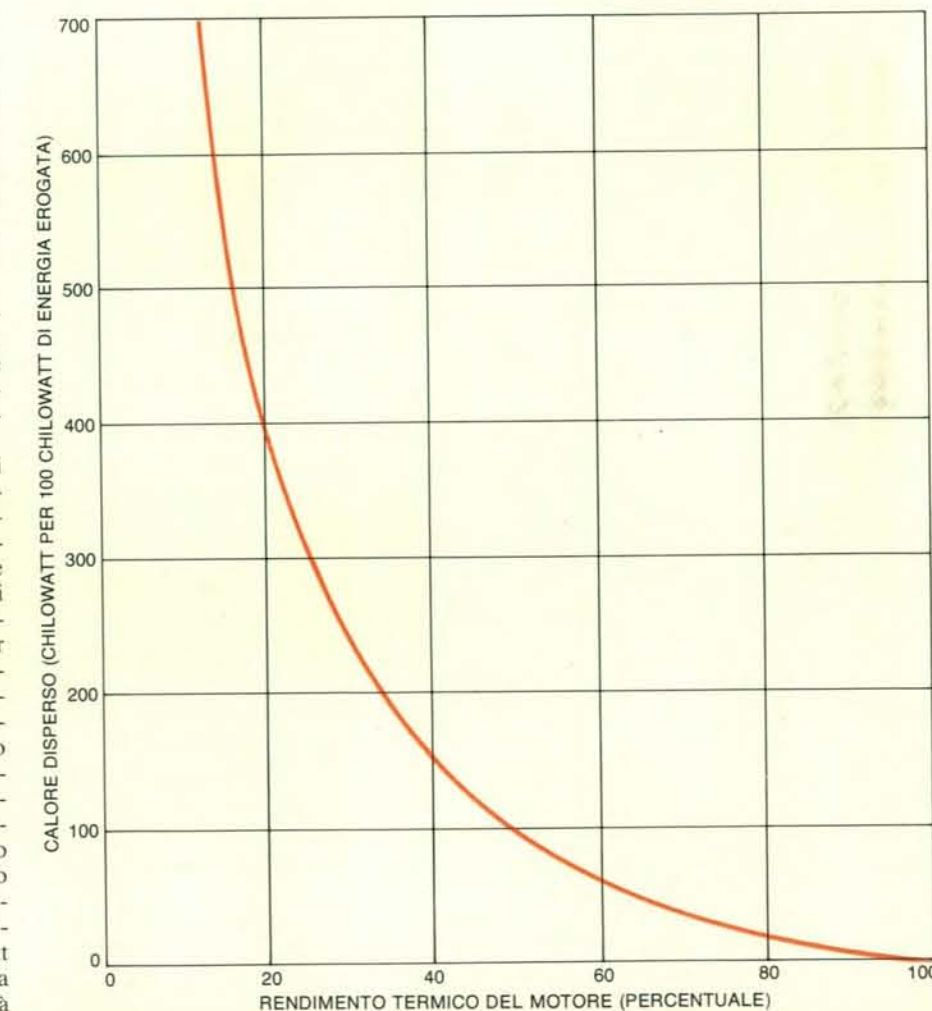
Dopo che il problema dell'inquinamento dell'aria causato dai motori a combustione interna venne collegato senza possibilità di dubbio alla presenza nei gas di scarico di idrocarburi, di ossido di carbonio e di vari ossidi di azoto (designati collettivamente con la formula NO_x), e dopo che risultò evidente che non sarebbe poi stato tanto facile ridurre tali emissioni, si levò un'ondata di entusiasmo per il motore a vapore come apparato motopropulsore alternativo, che avrebbe fatto scomparire tutti i problemi sollevati dalla composizione dei gas di scarico del motore a combustione interna. Ricordo personalmente uno di questi entusiasti, il quale mi parlava di un motore a vapore, grande più o meno quanto un pallone da football, capace di erogare una potenza di 75 chilowatt (100 cavalli vapore). Rimasi esterrefatto. Gli chiesi come avrebbe dovuto essere grande la caldaia, per alimentare un motore del genere, e quello mi rispose che essa era più o meno delle stesse dimensioni. Passai queste informazioni a un collega più anziano, il quale va sempre dritto al nocciolo della questione. «Prova a chiedergli - mi suggerì - quanto è grande il condensatore.» Quella era ovviamente la domanda chiave. Spiegherò perché.

Nel campo dei trasporti quasi tutta l'energia viene dai motori termici. Questi sono congegni che producono lavoro meccanico da un flusso di calore fornito da una fonte termica. Essi convertono una parte di questo calore in lavoro, e devono disperderne il resto in una «spugna» termica: l'ambiente circostante. I motori a combustione interna, le turbine a gas e via dicendo dipendono da un flusso di calore fornito da gas a temperatura elevata generati dalla combustione del carburante. Quasi tutte le turbine idrauliche (e tutte le navi a vela) ricavano la loro energia dai processi atmosferici attivati dalla differenza di temperatura che si ha quando la radiazione del Sole viene assorbita dalla Terra e dalla sua atmosfera a un certo livello energetico e quando essa viene irradiata indietro nello spazio a un livello energetico inferiore. Le uniche eccezioni all'applicazione universale dei motori termici a fini di trasporto comportano l'uso della forza muscolare. Un sistema fondamentalmente analogo di conversione diretta dell'energia, la pila a combustibile, è utile nelle lunghe missioni spaziali. Ma né la forza muscolare né la pila a combustibile avranno mai una parte significativa nei trasporti autostradali del futuro. Possiamo quindi concentrarci sui motori termici.

In qualsiasi motore termico il calore fluisce da una fonte a temperatura elevata in un dispositivo in grado di convertire una parte di tale calore in lavoro, disperdendone il resto in una «spugna» termica, la cui temperatura è più bassa di quella della fonte. Nel caso di un motore a stato stazionario, per la legge della conservazione dell'energia (la prima legge della termodinamica), l'energia che ne esce è uguale all'energia che vi entra. In altre parole, l'apporto di calore è uguale al lavoro più il calore disperso. Il rendimento termico di un motore può essere definito come la quantità di lavoro prodotto divisa per la quantità di calore immessa nel sistema. Per ridurre il consumo di energia nei veicoli a motore è necessario aumentare il rendimento termico del motore, soprattutto nelle condizioni di erogazione di potenza ridotta in cui il motore opera per la maggior parte della sua vita attiva.

Per un dato valore di rendimento, si può risparmiare energia nel campo dei trasporti solo riducendo il lavoro richiesto. Una buona parte dell'energia richiesta per dare propulsione a un veicolo è proporzionale alla massa del veicolo stesso. Si richiede energia, per esempio, per accelerare il veicolo dopo le fermate e i rallentamenti imposti dal traffico, per superare una salita e per vincere la resistenza di rotolamento dei pneumatici. La resistenza aerodinamica incomincia ad avere importanza solo quando si guida a velocità elevata per un periodo di tempo prolungato, e solo una piccola parte del carburante per automobili viene consumata in questo modo. Pertanto un requisito primario dei motori d'automobile, oltre all'elevato rendimento termico, è che essi siano leggeri, e un requisito secondario è che non accrescano la resistenza aerodinamica del veicolo.

Il processo di dispersione del calore si ripercuote, in alcuni motori, sia sulla massa sia sulla resistenza aerodinamica. Possiamo modificare l'equazione della conservazione dell'energia così da indicare che la dispersione di calore per unità di lavoro è una funzione inversa del rendimento del motore: il calore disperso per unità di lavoro prodotto è uguale a 1 diviso per il rendimento termico meno 1. L'equazione ci dice che, se abbiamo un motore di 100 chilowatt con un rendimento del 40 per cento, valore cui si può avvicinare un motore Diesel o un motore Stirling, si dovranno disperdere 150 chilowatt di calore. Se il rendimento è pari solo al 20 per cento, valore, questo, più tipico dei motori a vapore o a combustione interna di basso rendimento, bisognerà disperdere necessariamente 400 chilowatt di calore, per ottenere lo stesso lavoro. La necessità di disperdere calore in quantità così consistenti può esigere un prezzo molto alto. Essa significa, per esempio, che un motore a vapore deve portarsi dietro per forza un grosso condensatore attraverso il quale indurre un grande flusso d'aria.

FONTE DI CALORE A TEMPERATURA ELEVATA
CALORE IMMESSO
MOTORE
LAVORO
CALORE DISPERSO
SPUGNA TERMICA A BASSA TEMPERATURA

Secondo la prima legge della termodinamica, il flusso di energia che esce da un motore termico è uguale a quello che vi entra. Solo una parte della potenza erogata può tradursi in lavoro meccanico. Il resto viene disperso in una «spugna» termica. Il rendimento termico è definito dal lavoro utile diviso per la quantità di calore immessa nel sistema. Secondo il principio enunciato da Nicolas Léonard Sadi Carnot, il massimo rendimento termico raggiungibile da una qualsiasi macchina termica viene calcolato sottraendo da 1 il rapporto fra la temperatura assoluta alla quale il calore viene disperso e la temperatura assoluta alla quale il calore è immesso nella macchina.

La quantità di calore dispersa da un motore termico aumenta rapidamente a mano a mano che cala il rendimento del motore. Quello della dispersione del calore è un problema serio nei motori a ciclo chiuso caratterizzati da un rendimento moderato-basso, come i motori a vapore, perché il calore deve essere rimosso attraverso un condensatore, la cui presenza determina un aumento delle dimensioni, del peso e del costo del sistema. Nei motori a ciclo aperto, come il convenzionale motore a benzina a ciclo Otto, il calore di scarico viene disperso direttamente nell'atmosfera.

	RAPPORTO FRA LE TEMPERATURE	RENDIMENTO DEI COMPONENTI	RENDIMENTO IN RAPPORTO AL MASSIMO RENDIMENTO TERMODINAMICO	DISPOSITIVO TERMODINAMICO PER LA DISPERSIONE DEL CALORE
MOTORE OTTO	ALTO	DISCRETO	MEDIOCRE	NESSUNO
MOTORE DIESEL	ALTO	DISCRETO	MEDIOCRE	NESSUNO
MOTORE RANKINE	BASSO	BUONO	DISCRETO	GRANDE
MOTORE STIRLING	MODERATO	BUONO	MOLTO BUONO	GRANDEZZA MODERATA
MOTORE BRAYTON A CICLO APERTO (TURBINA A GAS)	MODERATO, POTENZIALMENTE ALTO	MOLTO BUONO	BUONO	NESSUNO
MOTORE BRAYTON A CICLO CHIUSO	MODERATO	MOLTO BUONO	BUONO	GRANDEZZA MODERATA

I motori termici adatti per le automobili presentano pregi e difetti di natura diversa. Il rapporto fra la temperatura all'ingresso e quella all'uscita determina il massimo rendimento termodinamico del sistema *(prima colonna)*. Le temperature più alte di immissione del calore si ottengono quando il carburante viene miscelato con aria e acceso in un motore Otto o Diesel. Il rendimento col quale l'energia liberata viene poi convertita in lavoro meccanico dipende dal rendimento dei vari componenti del motore *(seconda colonna)*. Un motore dotato di componenti efficienti si avvicina al massimo rendimento termodinamico più di un motore con componenti meno efficienti *(terza colonna)*. I motori a ciclo chiuso, in più, hanno bisogno di dispositivi per la dispersione del calore, quali gli scambiatori di calore *(ultima colonna)*.

La natura sembra punire i deboli e gli inefficienti e premiare i forti e gli efficienti. Motori come il motore a ciclo Otto, il motore Diesel e la turbina a gas a ciclo aperto non devono però pagare un prezzo troppo alto in termini di peso o di resistenza aerodinamica per la dispersione del calore, perché essi si servono del loro fluido operativo (in questo caso l'aria) per ossidare il carburante all'interno del motore. È per questo che vengono chiamati motori a combustione interna. Il fluido operativo perde energia e deve essere scaricato alla fine di ogni ciclo (donde il termine ciclo aperto), ed è necessario introdurre una nuova carica. Con i motori di questo genere il processo di dispersione del calore è compiuto pertanto dall'aria. (Il calore che viene disperso nel radiatore di un motore a pistone rappresenta un problema meccanico più che termodinamico: l'olio lubrificante e alcuni componenti critici devono essere mantenuti a temperature ragionevolmente basse. Il rendimento del motore migliorerebbe se fosse possibile adoperare materiali e lubrificanti che non esigono il raffreddamento del cilindro.)

Per riassumere in breve la tesi fin qui esposta, si può dire che, se si raggiungesse un grado elevato di rendimento in ogni tipo di motore per autoveicoli, oltre all'ovvio beneficio del risparmio di carburante, se ne avrebbe un altro, meno ovvio, che riguarda solo i motori a combustione esterna a ciclo chiuso, poiché la quantità di calore che deve essere dispersa si riduce in misura notevole a mano a mano che aumenta il rendimento termico. In un motore a ciclo chiuso, come per esempio un motore a vapore a condensazione, il fluido operativo (che nel caso del motore a vapore è l'acqua) è chiuso ermeticamente nel motore e viene riscaldato da una combustione esterna al fluido stesso. Ecco perché il fluido operativo qui deve essere raffreddato in un dispositivo apposito e non scaricato a temperatura elevata nell'atmosfera, come invece avviene nei motori a ciclo aperto.

Un corollario della seconda legge della termodinamica è che il massimo rendimento possibile di un motore termico dipende soltanto dal rapporto fra la temperatura alla quale il calore viene fornito e quella a cui viene disperso. (La semplice equazione è: il massimo rendimento termico è uguale a 1 meno 1 diviso per il rapporto fra quelle due temperature.) Se in un motore si vuole raggiungere un grado elevato di rendimento, bisogna che sia alto anche il rapporto fra le due temperature. Questa condizione, però, è necessaria ma non sufficiente. Un motore con un rapporto fra le temperature piccolo, non può avere un grado elevato di rendimento. Un motore con un alto rapporto di temperatura avrà un grado elevato di rendimento termico solo se si avvale di processi interni che non introducono nessuna perdita termodinamica fra le tante possibili.

Confrontiamo fra loro le fasi operative e alcune parti di motori termici tipici. Tutti si basano su quattro processi, o fasi operative, fondamentali: pompaggio o compressione, apporto di calore, espansione, dispersione del calore. Nei motori a combustione interna le prime tre fasi operative si svolgono in ordine successivo in un'unica parte. Trattandosi di motori a ciclo aperto, la quarta fase, cioè la dispersione del calore, avviene nell'atmosfera.

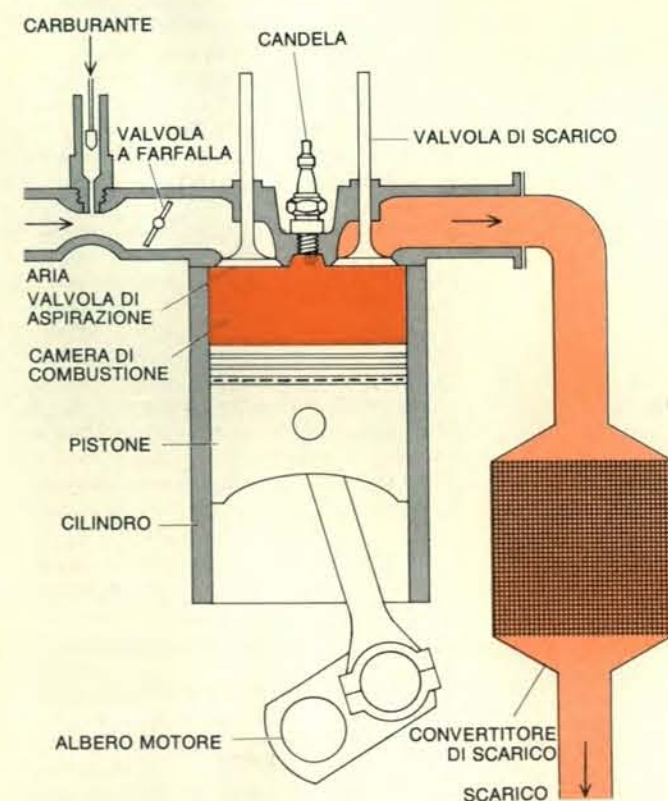
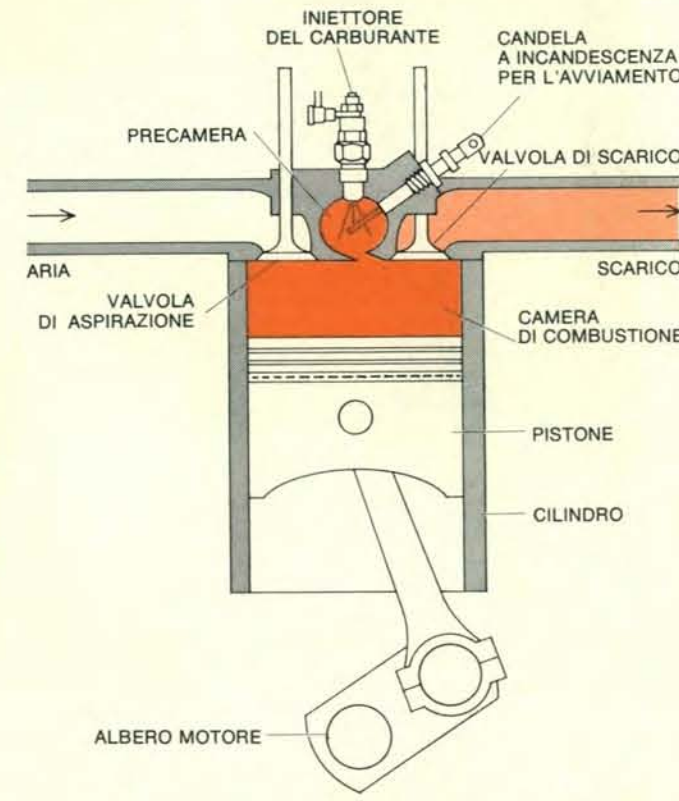
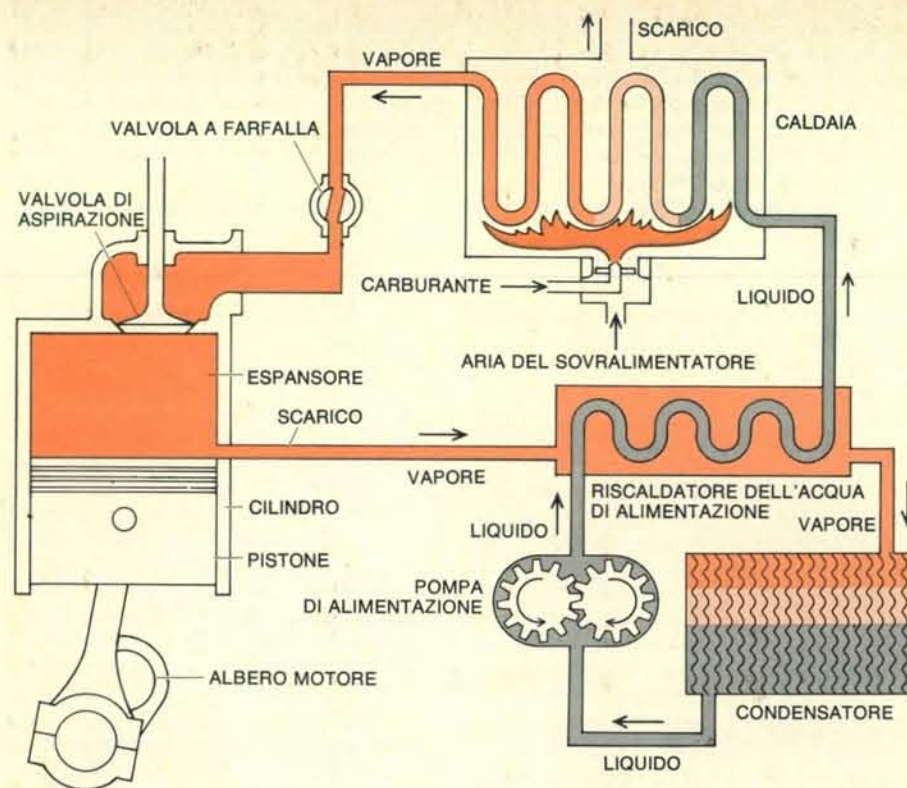
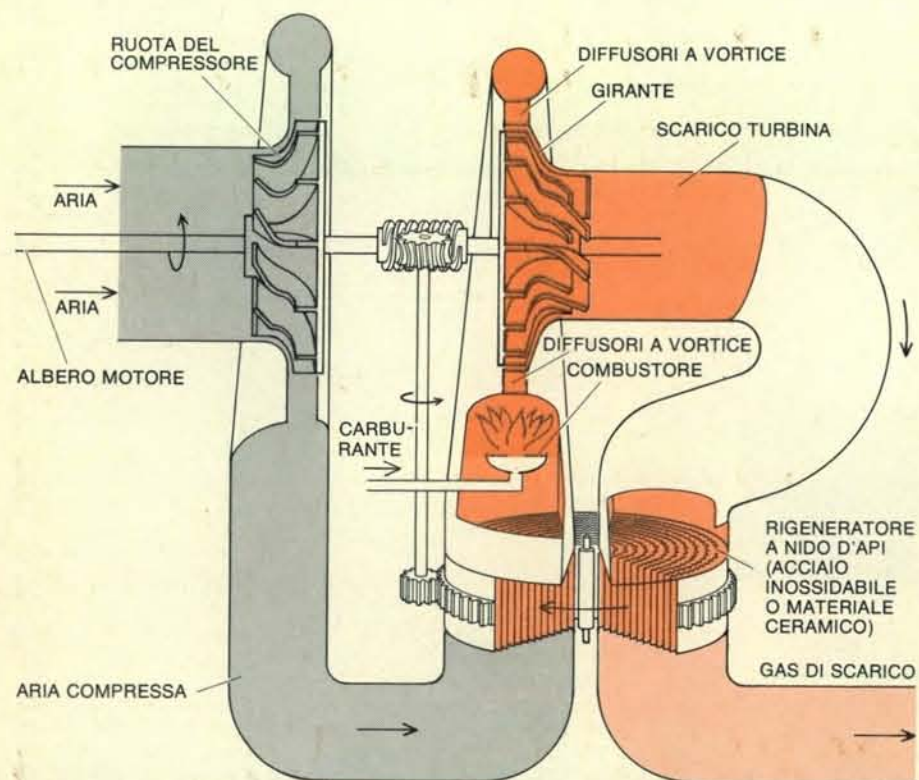
Nel motore a ciclo di vapore (che può funzionare o col vapore o con fluidi meno familiari, come il mercurio o il freon 12) e nella turbina a gas, le quattro fasi operative si svolgono ininterrottamente in parti separate. Una di queste è impegnata costantemente a pompare o comprimere, un'altra a riscaldare il fluido e una terza a far espandere il fluido stesso. Nel caso della turbina a gas a ciclo chiuso e della turbina a vapore, c'è anche una quarta parte in attività costante, e cioè il refrigerante o condensatore. Queste parti possono essere perfezionate per ogni singola fase, in modo che il loro rendimento sia più elevato di quanto lo sia per le fasi che si svolgono in successione in un'unica parte (il cilindro) di un motore a combustione interna. Alcuni motori hanno anche un rigeneratore, o scambiatore di calore, il quale ha la funzione di trasferire il calore dal fluido operativo caldissimo a bassa pressione al fluido che si trova nella parte del ciclo più fredda e ad alta pressione.

Il motore Stirling, nella forma odierna tipica, è un motore a pistone e cilindro a combustione esterna, che come fluido operativo usa una piccola quantità di idrogeno, contenuta in uno spazio chiuso ermeticamente. Inventato nel 1816 da Robert Stirling, un pastore scozzese, questo motore è un congegno di notevole raffinatezza termodinamica e fu la prima macchina termica a sfruttare il principio della rigenerazione. Il suo funzionamento potrà essere compreso meglio con l'ausilio della figura alle pagine 14 e 15. Ma, ai fini del nostro discorso, non è necessario soffermarsi sui dettagli.

La divisione dei motori in due categorie: quelli cioè in cui le varie fasi operative si svolgono in successione in un'unica parte, e quelli in cui le quattro fasi si svolgono senza soluzione di continuità in parti separate, ha una relazione importante con il massimo rendimento termico potenziale. La successione delle fasi in un motore a pistoni a combustione interna è così rapida, che si possono tollerare dei gas a temperature elevatissime (dell'ordine di 2500 gradi centigradi). Le pareti della camera di combustione rimangono a una temperatura più bassa, tale da potersi conciliare senza pericolo con parti di ghisa o di alluminio e con un olio lubrificante abbastanza comune.

Nei motori in cui le fasi operative si svolgono senza soluzione di continuità, invece, le pareti della camera di combustione e dell'apertura di entrata dell'espansore sono continuamente a contatto con gas a temperatura elevata. La massima temperatura consentita rimane così limitata a quella alla quale possono resistere i materiali dei quali si può disporre (o che ci si può permettere). Nei motori a combustione esterna a ciclo chiuso, come il motore a vapore, la turbina a gas a ciclo chiuso e il motore Stirling, il calore immesso nel sistema deve attraversare le pareti del dispositivo di riscaldamento (per esempio i tubi della caldaia). Queste pareti pertanto si trovano a una temperatura più elevata della temperatura massima del fluido operativo e il deterioramento termico dei materiali delle pareti diventa così un limite assoluto. Nei motori a combustione interna e a ciclo aperto in cui le fasi operative si svolgono senza soluzione di continuità, come le turbine a gas a ciclo aperto, la temperatura del gas può essere parecchie centinaia di gradi al di sopra della temperatura massima consentita dalle pareti, perché il calore viene liberato non dalla parte esterna di una parete, bensì all'interno del flusso.

Molti sforzi vengono dedicati attualmente al tentativo di raffreddare le pareti di quelle parti delle turbine a gas in cui avviene l'immissione del calore e l'espansione, in modo da poter sfruttare per i gas temperature ancora più elevate. Si stanno inoltre sperimentando, nella costruzione delle turbine a gas, materiali ceramici, non solo per le parti stazionarie, quali i combustori e le pale dell'ugello, ma anche per la girante. Oggi negli aeroplani commerciali di tipo avanzato si hanno, all'entrata della turbina, temperature che vanno dai 1250 ai 1350 gradi centigradi, e negli apparecchi militari le temperature sono almeno 100 o 200 gradi in più. Sui banchi di prova si raggiungono, da almeno un decennio, temperature molto superiori ai 1650 gradi centigradi. Le turbine a gas per autotrasporto usano l'atmosfera per la dispersione del calore, di modo che la temperatura effettiva della dispersione stessa è di circa 25 gradi centigradi. Il rapporto fra la temperatura (assoluta, cioè in gradi sopra lo zero assoluto) alla quale il calore viene immesso nel sistema e quella alla quale viene scaricato nell'atmosfera dovrebbe essere almeno di cinque a uno, in modo da rendere possibile ai motori di raggiungere un rendimento termico elevatissimo quando sono a pieno carico.

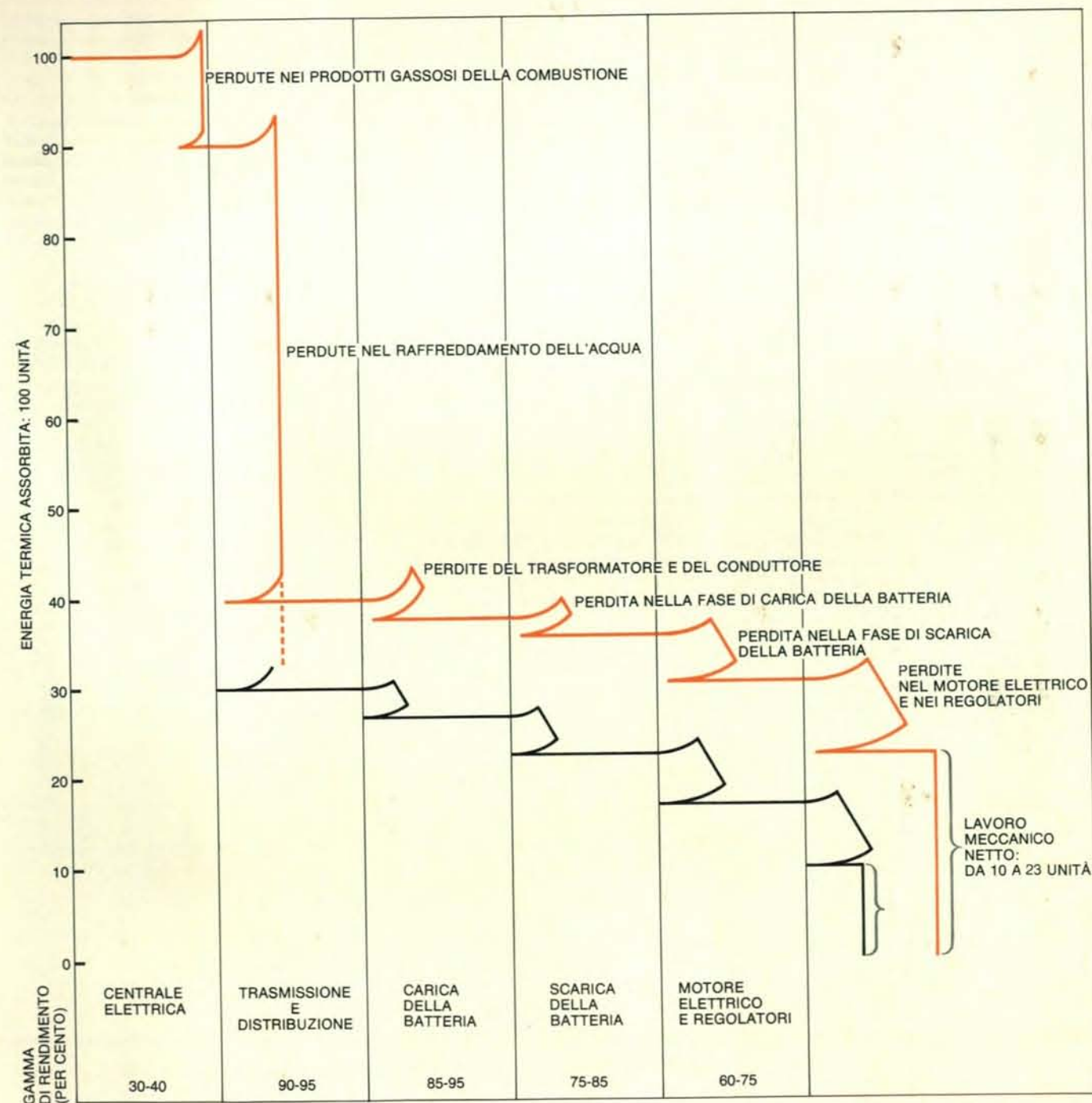
Purtroppo i motori a ciclo di vapore non possono usare all'entrata temperature di quest'ordine di grandezza, non solo perché il calore deve essere trasferito attraverso le pareti di un dispositivo di riscaldamento che sono sottoposte a un'estrema sollecitazione, ma anche perché i possibili fluidi operativi hanno, a temperature elevate, particolari proprietà. Al di sopra dei 600 gradi centigradi, per esempio, il vapore diventa quasi corrosivo, facendo sì che gli acciai assorbano idrogeno e diventino fragili. Ormai da parecchi decenni nelle turbine a vapore dei gruppi motori stazionari è stata quasi universalmente adottata una temperatura massima di 566 gradi centigradi, in seguito a disastrose avarie verificatesi in turbine che operavano a temperature più elevate. Alcuni motori a turbina sperimentali per automobili hanno operato però a temperature notevolmente al di sopra dei 566 gradi, grazie all'uso di materiali piuttosto costosi. I migliori fluidi organici che siano stati utilizzati per i motori a ciclo di

Nel motore convenzionale (a ciclo Otto) una candela viene utilizzata per causare l'accensione di una miscela di carburante e aria. L'energia contenuta nei gas che si espandono viene assorbita da un pistone e trasmessa quindi a un albero a gomiti. Nella sua corsa di ritorno, il pistone scarica i gas bruciati, preparando il cilindro per la carica successiva di carburante. Negli USA quasi tutte le nuove automobili sono state munite di convertitori catalitici per ridurre le emissioni di scarico.

Il motore Diesel fa a meno della candela e accende la miscela del carburante nel cilindro mediante il calore della compressione. Il rapporto di compressione di un motore Diesel tipico è di circa 20 a 1, vale a dire più del doppio di quello della maggior parte dei motori con accensione a scintilla. Una candela a incandescenza facilita la messa in moto. Fatta eccezione per gli ossidi di azoto (NO_x), i motori Diesel liberano sostanze inquinanti in quantità inferiore rispetto ai motori a ciclo Otto.

Il motore a vapore a ciclo Rankine trasmette il calore prodotto dalla combustione del combustibile a un sistema chiuso, contenente un fluido operativo condensabile, che nel caso dei motori a vapore è l'acqua. Il fluido operativo viene vaporizzato e surriscaldato in una caldaia, dopo di che viene fatto espandere per produrre potenza e quindi condensare per poterlo riutilizzare. Nel caso delle automobili, alle turbine si preferiscono i pistoni, i quali producono spostamenti meccanici.

La turbina a gas (motore Brayton) è azionata dall'espansione di gas caldissimi attraverso una girante. Una parte della potenza prodotta dalla turbina viene utilizzata per comprimere l'aria che viene immessa nella camera di combustione del motore. Nelle forme che sono attualmente in fase di sviluppo per le automobili, i gas di scarico della turbina passano attraverso un rigeneratore, o ricuperatore di calore, a nido d'ape, il quale ha la funzione di riscaldare l'aria in arrivo.

vapore si decompongono quando vengono riscaldati oltre i 400 gradi circa. I vapori di mercurio furono adoperati negli anni trenta a una temperatura massima di circa 650 gradi in alcuni cicli a due fluidi e più di recente a una temperatura di 705 gradi in alcuni elementi di veicoli spaziali. Il mercurio pone però dei problemi straordinari, fra cui, per esempio, quello della bagnatura delle superfici del condensatore per poter raggiungere un grado di rendimento elevato.

La natura quindi non è stata benigna con i cicli a vapore. Essa ha posto dei limiti alla temperatura alla quale si può immettere calore nel sistema, riducendo in tal modo il rendimento massimo potenziale, il che a sua volta fa aumentare enormemente la quantità di calore che si deve necessariamente disperdere. Tuttavia nei grandi gruppi motori stazionari si possono raggiungere rendimenti termodinamici pari a qualcosa come il 44 per cento, riducendo al minimo le perdite termodinamiche e raggiungendo la più bassa temperatura possibile per la dispersione del calore, nonostante il limite ridotto posto alla temperatura di entrata del calore stesso. Il condensatore, che oltre tutto costa parecchio, non deve essere troppo grande finché si può disporre di acqua fredda di mare o di fiume. Altrimenti, per disperdere il calore, si dovranno installare enormi torri di raffreddamento. Le attuali centrali termonucleari sono costrette a usare per il vapore, all'entrata, temperature ancora più basse di quelle adottate nelle centrali alimentate da combustibile fossile, col risultato che nelle centrali termonucleari i costi di dispersione del calore sono molto alti, anche qualora si abbia la possibilità di raffreddare i condensatori con acqua di mare o di fiume.

Nei motori di piccole dimensioni richiesti per i veicoli non è possibile fare ricorso alle misure speciali, come quella di aumentare il numero di fasi della turbina o di riscaldare di nuovo il vapore durante l'espansione e di riscaldare l'acqua di alimentazione prelevando un po' di vapore dalla turbina, adottate nei grandi impianti motori stazionari per ridurre le perdite termodinamiche. Ne consegue che il massimo rendimento possibile cala fino a circa il 30 per cento. Anche questo valore implica che la temperatura di dispersione del calore si avvicini a quella possibile nei grandi impianti di potenza, si avvicini cioè a una temperatura dell'ordine di circa 35-40 gradi centigradi. Per raggiungere questo livello con una temperatura moderata dell'aria di 25 gradi centigradi ci vorrebbe però un flusso d'aria estremamente abbondante e un condensatore tanto grande da superare in dimensioni qualunque veicolo.

Perciò il progettista di motori d'automobile rende più elastici i requisiti imposti ai flussi d'aria, permettendo un aumento più sensibile della temperatura dell'aria, il che significa che bisogna aumentare la temperatura di condensazione, che il rendimento termico si deteriora ulteriormente, e che aumenta ancora di più la quantità di calore da disperdere. Il condensatore però non diventa più piccolo. Un condensatore di dimensioni accettabili, quasi vicine al minimo, è possibile di solito a una temperatura all'uscita del calore compresa fra i 95 e i 120 gradi centigradi, intervallo nel quale il rendimento termico globale, tenuto conto della potenza necessaria per azionare la ventola del condensatore e un'altra ventola per spingere l'aria nella camera di combustione, scende, per gli attuali cicli a vapore, al 22-24 per cento.

Riduzioni marginali nelle dimensioni del condensatore si possono raggiungere con alette migliori sulle superfici tubolari o piatte del condensatore stesso. Qualunque riduzione sostanziale richiederebbe però o una nuova età glaciale, che renda possibile avere temperature più basse di dispersione del calore con un rendimento più elevato e quantità più ridotte di calore da disperdere, o temperature del vapore

Il rendimento della propulsione elettrica per le automobili dipende in misura notevole dal rendimento col quale le unità di calore del combustibile originale vengono convertite in elettricità alla centrale elettrica e poi trasmesse al consumatore, prima che questi possa incominciare a caricare le batterie nell'automobile stessa. La curva in colore rappresenta più o meno il meglio che si possa ottenere in ogni fase con le tecnologie note. Su 100 unità di energia, soltanto 38 raggiungono l'apparecchiatura per la carica delle batterie, e soltanto 23 emergono alla fine sotto forma di energia meccanica per la propulsione del veicolo. Nel caso peggiore, in cui per ogni tappa del processo il rendimento è quello minimo indicato in figura, solo 10 unità si ritroverebbero, alla fine, sotto forma di energia meccanica. Il grafico può essere fuorviante, e indurre a concludere che le perdite del condensatore sono le più significative. In realtà, quelle del condensatore sono perdite di calore a bassa temperatura, il quale ha un potenziale ridotto di produzione di lavoro. Le perdite di gas della combustione riguardano calore a temperatura un poco più elevata, e sono quindi una perdita più consistente; le altre sono perdite di energia molto pregiata, quindi assai più costose dal punto di vista termodinamico. Il grafico si basa su un'analisi compiuta da Paul D. Agarwal della General Motors Corporation.

molto più alte all'entrata dell'espansore, attraverso lo sfruttamento di materiali che non siano soggetti a essere resi fragili dall'idrogeno.

Ho parlato fin qui dei limiti di rendimento termico dei gruppi motori solo dal punto di vista del rapporto globale fra l'immissione e la dispersione del calore, e dell'ulteriore aggravio, imposto ai motori a ciclo chiuso, di dover elevare la temperatura al momento della dispersione del calore, quando il rendimento del motore scende al di sotto di un certo valore critico, che corrisponde grosso modo al 30 per cento. Il rendimento delle singole parti del motore, il rendimento cioè di ogni singola fase operativa, può variare moltissimo da un motore all'altro e da un ciclo all'altro. Nel caso del motore a vapore, la compressione, che è la prima delle fasi operative fondamentali che si svolgono in successione in tutti i motori termici, assorbe soltanto una piccola parte di tutto il lavoro utile del motore, perché il fluido che viene pompato (immesso a forza nella caldaia) è allo stato liquido. Non ha perciò nessuna importanza il fatto che la pompa abbia un rendimento pari al 90 o soltanto al 75 per cento: la quantità di «lavoro negativo» che la pompa è chiamata a svolgere è molto ridotta.

Completamente diverso è il caso della turbina a gas, dove il fluido che viene compresso è un gas, e la quantità di lavoro negativo è molto grande. Un personaggio di Charles Dickens, Mr. Micawber, definisce la miseria come la condizione caratterizzata dal fatto di avere un reddito inferiore del 5 per cento alle spese necessarie, e la felicità, per contro, come la condizione caratterizzata da un reddito superiore del 5 per cento alle spese in questione. I progetti delle prime turbine a gas erano manchevoli, perché i compressori (le pompe) assorbivano più potenza di quanta ne producessero i dispositivi di espansione. Il lavoro utile netto di una turbina a gas è dato dalla differenza fra queste due grandi e spesso analoghe quantità. Per questa ragione io chiamo la turbina a gas «motore di Micawber», un motore cioè molto sensibile al rendimento dei componenti.

La prima turbina a gas con la quale ho avuto a che fare era capace soltanto di sostenere se stessa, e non era in grado di produrre il benché minimo lavoro utile. Date le circostanze, un miglioramento dell'1 per cento del rendimento dell'espansore o del compressore avrebbe aumentato il rendimento globale praticamente da zero a qualche valore limitato, rendendo così il motore infinitamente sensibile al rendimento dei suoi componenti. Un modo per aumentare il lavoro utile dell'espansore di una turbina a gas è ovviamente quello di aumentare la temperatura del gas all'entrata. Un aumento del genere accrescerà anche il rendimento termico globale del motore, a patto che rimanga inalterata la temperatura che si registra all'uscita.

A mano a mano che le turbine a gas accrescono il proprio rendimento grazie all'aumento della temperatura all'entrata dell'espansore, aumenta la differenza fra lavoro utile dell'espansore e il fabbisogno di energia del compressore, e diminuisce la sensibilità del rendimento termodinamico del motore nei confronti del rendimento delle sue parti. Inoltre le varie parti sono state progettate in modo da raggiungere un alto grado di rendimento individuale.

Un motore a pistoni con accensione a scintilla cela dentro di sé un dispositivo di espansione che produce una potenza molto superiore a quella dei cavalli vapore del motore dichiarati ufficialmente, e un compressore (il pistone e il cilindro stessi) che assorbe gran parte della differenza fra la potenza del dispositivo di espansione e il lavoro effettivo del motore. Dato però che possono tollerare temperature elevate, i motori a pistoni non sono molto sensibili al rendimento delle proprie fasi operative separate. Ma la cosa non ha nessuna importanza, perché i compromessi, a cui si è dovuto ricorrere nella fabbricazione di un solo componente - una combinazione di cilindro e pistone - per fare sì che il componente in questione serva a tre funzioni, hanno l'effetto di ridurre in maniera significativa il rendimento delle singole fasi operative.

Il rendimento termico di un motore che opera a pieno carico però non è tutto, quando si confrontano fra loro vari tipi di motore. Una parte dei motori d'automobile (soprattutto in paesi come gli USA in cui vigono limiti di velocità molto restrittivi) passano la maggior parte della loro vita attiva operando a potenze inferiori al 25 per cento di quella massima. La turbina a gas convenzionale, in particolare, subisce un brusco calo di rendimento a carichi parziali. Per ridurre la potenza in una normale turbina a gas a ciclo aperto, si riduce il flusso del combustibile, il che riduce sia il rapporto globale tra le temperature sia la pressione del gas che lascia il compressore. Il rendimento del compressore può subire un calo netto rispetto al piano di progettazione, specie se il compressore stesso è progettato per rapporti di pressione elevati. Pertanto le turbine a gas a ciclo aperto con un rapporto alto di pressione hanno un rendimento mediocre a carico parziale. C'è però la possibilità, per quel che riguarda le turbine a gas, di costruirle in molte forme alternative, alcune delle quali hanno un grado di rendimento accettabile a carico parziale.

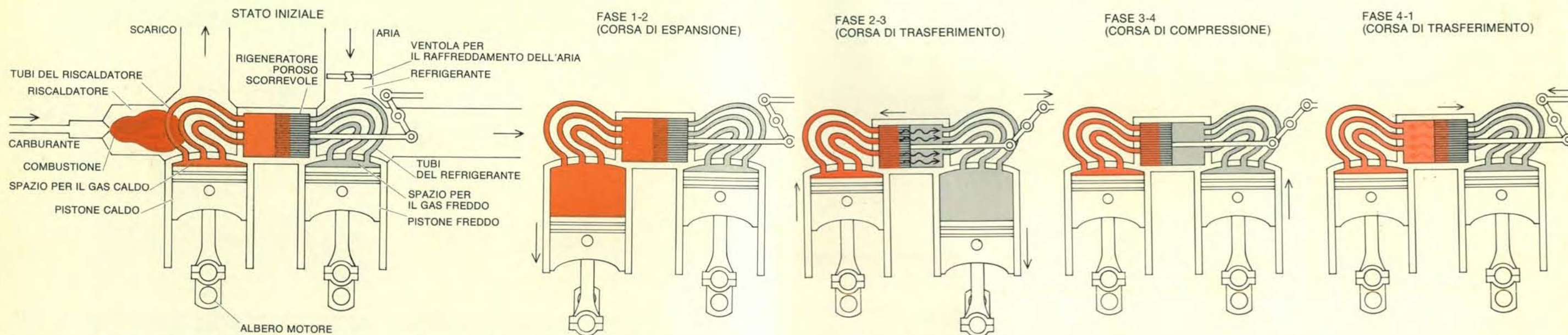
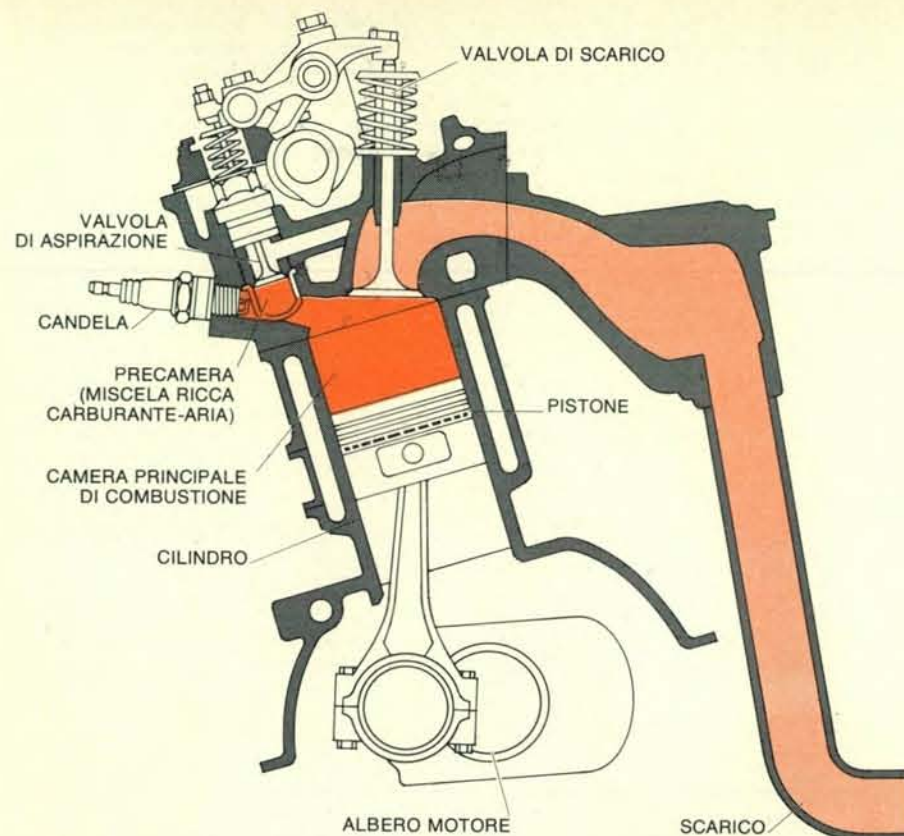
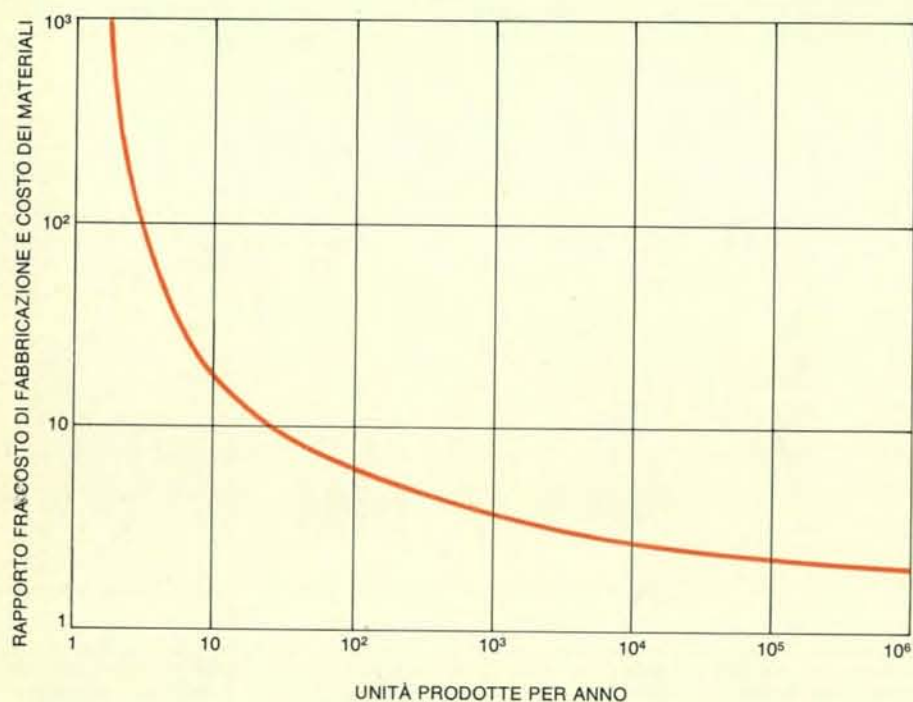
Tutti gli altri motori alternativi possibili si comportano abbastanza bene a basso carico dal punto di vista del rendimento. Il rendimento dei motori Diesel e Otto diminuisce, sì, a basso carico, ma non in una misura inaccettabile. Il rendimento dei motori Stirling e di alcuni tipi di motori a ciclo di vapore e di motori a turbina a gas a ciclo Rankine può invece aumentare a basso carico. Questo è dovuto in parte al fatto che gli scambiatori di calore normalmente accrescono il proprio rendimento quando trattano flussi di massa più ridotta a velocità più basse, in parte al fatto che in genere i componenti presentano perdite di pressione più limitate quando viene ridotta la velocità, e in parte al fatto che a basso carico è possibile eliminare le perdite parassitarie delle ventole per il condensatore o refrigerante.

L'impulso iniziale che portò alla ricerca di un motore nuovo per le automobili costituiva un riconoscimento delle grandi quantità di sostanze inquinanti che venivano emesse dai motori di solo un decennio fa. Nel motore convenzionale, molte di queste sostanze si formano in seguito al processo di combustione intermittente, col suo rapido raffreddamento dei prodotti della combustione stessa. Tutti i motori a combustione continua emettono sostanze inquinanti in quantità molto più ridotta di quanto faccia, in assenza di meccanismi regolatori, il motore con accensione a scintilla. Modelli sperimentali di motori a ciclo Rankine, di motori a turbina a gas e di motori Stirling hanno superato le norme più severe fin qui stabilite in materia di emissioni di scarico.

Pur essendo anch'esso a combustione intermittente, il motore Diesel ha il vantaggio di operare con aria in eccesso, per cui la sua emissione di ossido di carbonio è trascurabile. Inoltre gli idrocarburi costituiscono solo una piccola parte degli scarichi del motore Diesel. Gli ossidi di azoto tuttavia si formano in fiamme caldissime e vengono «congelati» da un rapido raffreddamento. Per il momento il motore Diesel non può soddisfare le norme severe sugli NO_x imposte per il 1980 e oltre. Vi sono anche dei sospetti, di cui però non è stata accertata la validità, che i gas di scarico dei motori Diesel contengano sostanze cangerogene.

Data la difficoltà di soddisfare completamente in un unico motore tutti i requisiti relativi al rendimento e alle emissioni di scarico, vi sono sistemi di propulsione ibridi che hanno i loro patrocinatori. Da parte mia, verrà definito ibrido un motore composto da un piccolo motore a regime costante unito a qualche forma di immagazzinamento di energia destinata a fornire una potenza supplementare, tale da soddisfare le varie normali esigenze di carico. Questi ibridi paiono pieni di promesse, perché i grandi motori convenzionali con accensione a scintilla hanno un rendimento bassissimo e sono nelle condizioni peggiori nel generare sostanze inquinanti quando funzionano a bassa potenza. Di conseguenza la maggior parte dei motori ibridi concepiti comporta un piccolo motore a regime costante, preferibilmente Diesel o con accensione a scintilla, di cui è facile regolare le emissioni, unito a un dispositivo di immagazzinamento di energia a volano, a batteria o ad aria compressa.

Ora, mentre sulla carta spesso appaiono buoni, questi sistemi di solito risultano pesanti, complessi e costosi. Possono certo essere progettati in modo da muovere in maniera soddisfacente un'automobile in condizioni normali, ma non possono uguagliare le prestazioni delle auto attuali nei lunghi tratti in salita o in altre circostanze che richiedono dal motore uno sforzo forte e prolungato, come, per esempio, quando la macchina traina una roulotte.

Infine dobbiamo considerare il caso del veicolo a trazione elettrica integrale, che non scarica gas di nessun genere, per lo meno in fase di esercizio, se non al momento in cui l'energia viene erogata originariamente. Le automobili e gli autocarri leggeri azionati da batterie di accumulatori e di motori elettrici hanno due svantaggi notevoli. Il primo è che, rispetto agli attuali motori a benzina, tanto le batterie di accumulatori quanto i motori elettrici sono molto pesanti per le rispettive funzioni, vale a dire l'immagazzinamento di energia e la generazione di potenza. Sarà possibile, certo, produrre batterie molto più leggere e durevoli di quelle odierne, ma sembra ancora probabile che, assieme, la massa formata dalle batterie, dal motore elettrico e dal meccani-

Il motore Stirling, inventato nel 1816 da Robert Stirling, è stato il primo motore termico a valersi del principio della rigenerazione, che rende possibile ricuperare energia termica che altrimenti andrebbe perduta. Il motore Stirling ha molte forme possibili; quella schematica (e ipotetica) della figura è stata scelta per chiarire il principio del suo funzionamento. Il calore della combustione viene assorbito ininterrottamente da un gas, che di solito è idrogeno, contenuto in un sistema chiuso ermeticamente. Allo stato iniziale lo spazio destinato al gas viene riempito di gas caldissimo a pressione massima. Entrambi i i pistoni si trovano in alto. Il rigeneratore è nella parte destra, o fredda, del motore. Nella corsa di espansione, fase 1-2, il pistone caldo viene spinto in giù dal gas che si espande ad alta pressione, erogando energia meccanica a un volano e a un albero motore. I tubi del riscaldatore continuano a riscaldare il gas che si espande. Nella corsa di trasferimento, fase 2-3, il pistone caldo e quello freddo si muovono in senso opposto, mentre il rigeneratore si sposta verso sinistra. In pratica tutto il gas viene costretto a passare attraverso il rigeneratore, il quale «assorbe» e immagazzina una buona parte del calore contenuto nel gas stesso. Siccome il gas diminuisce di temperatura a volume costante, la sua pressione scende. Nella corsa di compressione, fase 3-4, il pistone freddo viene spinto in su, comprimendo il gas, mentre la sua temperatura è mantenuta costante dai tubi del refrigerante. Nella seconda corsa di trasferimento, fase 4-1, i pistoni rimangono nella loro posizione più elevata, mentre il rigeneratore si muove verso destra. A mano a mano che passa attraverso il rigeneratore, il gas raccoglie il calore immagazzinato da quest'ultimo, aumentando quindi di temperatura a volume costante. Pertanto la pressione del gas aumenta, creando le condizioni per la successiva corsa di espansione e la ripetizione del ciclo.

Il motore a carica stratificata soddisfa le norme relative agli scarichi senza richiedere un sistema di ricircolazione dei gas di scarico o un convertitore catalitico. L'esemplare in figura, l'Honda CVCC, ha una camera principale e una precamera per ogni cilindro. Una miscela ricca carburante-aria entra nella precamera nel medesimo tempo in cui una miscela molto povera entra nella camera principale attraverso una valvola di aspirazione separata (non visibile nella figura perché situata dietro la valvola di scarico). La candela accende la miscela ricca, che a sua volta accende la miscela principale, quella povera. Così si ha combustione lenta e costante. Le temperature massime sono abbastanza basse da ridurre al minimo la formazione di ossidi d'azoto, e la temperatura media è abbastanza alta da limitare la liberazione di ossido di carbonio e idrocarburi.

I costi di fabbricazione di un motore sono legati al costo dei materiali, e calano sensibilmente con l'aumentare del volume. In effetti il costo dei materiali determina quello totale del motore: rispetto a quelli dei modelli alternativi i materiali delle vetture attuali sono molto economici.

smo di regolazione debbano superare, in chilogrammi per chilowatt, il valore che può essere raggiunto dai sistemi di trazione alimentati a carburante. Un simile handicap in fatto di peso comporta un correlativo handicap in fatto di perdita di energia, anche a parità di rendimento dell'apparato motopropulsore.

Il secondo e forse cruciale svantaggio è il fatto che il rendimento globale della propulsione elettrica, misurato dal consumo di combustibile della centrale elettrica fino all'energia erogata sulla strada, probabilmente non supera il rendimento possibile con qualunque altro sistema alternativo di propulsione, e anzi potrebbe essere forse parecchio inferiore. Per calcolare il rendimento dell'energia erogata alle ruote di un veicolo elettrico, bisogna moltiplicare in ordine il rendimento di cinque fasi operative separate: il rendimento termico della centrale elettrica, il rendimento della rete che eroga l'energia ai singoli utenti, il rendimento del caricatore per accumulatori, il rendimento del ciclo di carica e ricarica della batteria e il rendimento del motore elettrico e del combinatore. Se si prende il più alto rendimento plausibile per ogni elemento della catena, il prodotto sequenziale è $0,40 \times 0,95 \times 0,95 \times 0,85 \times 0,75$, ovvero il 23 per cento. In realtà (almeno per quel che riguarda gli Stati Uniti) il rendimento medio delle centrali termiche alimentate da combustibili fossili è pari soltanto al 33 per cento, e il rendimento termico delle centrali termonucleari è inferiore di circa l'uno per cento. Se il prodotto delle cinque fasi successive viene calcolato per una serie di valori più tipica, il rendimento netto della propulsione elettrica integrale scende a circa il 12 per cento.

Il rendimento termico, va da sé, non è l'unica misura della validità di una soluzione. I combustibili liquidi sono una merce sempre più scarsa e costosa. I veicoli a batteria elettrica possono fare uso di energia ricavata dal carbone o da combustibili nucleari (anche se per taluni quest'ultima alternativa è tutt'altro che allettante). Inoltre le batterie si potranno caricare in massima parte durante la notte, contribuendo a migliorare la distribuzione del carico (e quindi anche il rendimento) delle centrali erogative di potenza. Se tali vantaggi saranno o meno sufficienti a compensare gli svantaggi della propulsione elettrica è cosa, per il momento, non ancora chiara. In ogni caso i confronti che ho fatto riguardano soltanto veicoli che si possono sostituire direttamente all'automobile di oggi, adatta a ogni uso. Le automobili a batteria elettrica, adatte soltanto per brevi percorsi locali da compiersi a velocità ridotta, come andare e tornare dal lavoro e andare a far spese, richiederebbero batterie più piccole, imporrebbero un handicap minore in fatto di peso, e sarebbero quindi più desiderabili per tali scopi.

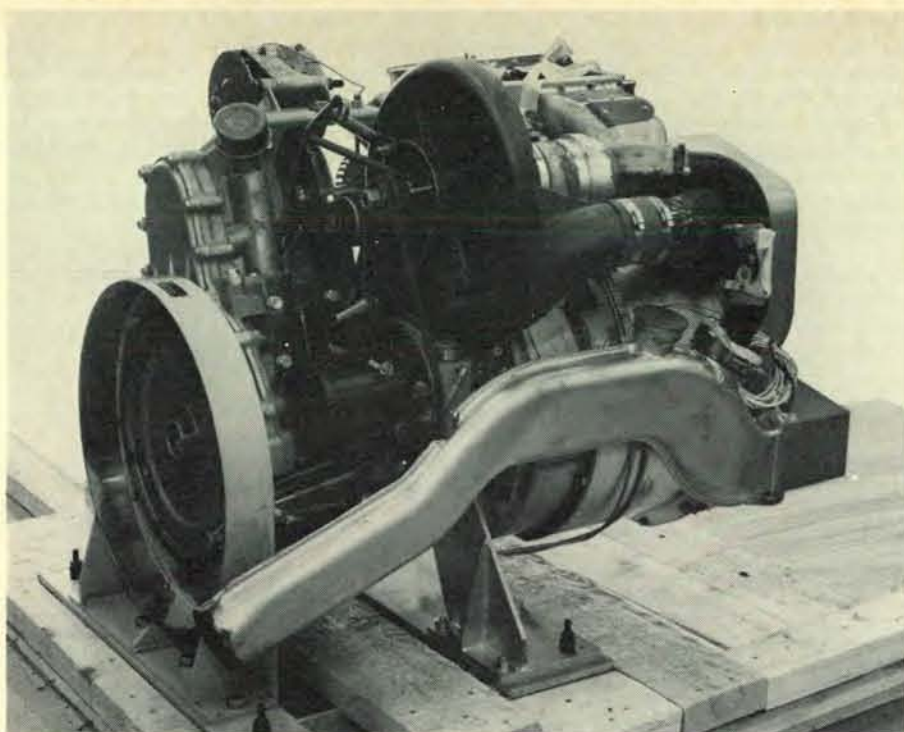
Un fattore critico che spesso viene trascurato nel valutare i sistemi di propulsione elettrica integrale è l'enorme carico che si verrebbe a riversare sull'industria elettrica, se una percentuale consistente di tutti i veicoli che circolano sulle strade fosse dotata di trazione elettrica. Infatti, nel caso degli USA, sarebbe necessario più o meno un quinto di tutta la nuova capacità generativa in programma da qui al 1990 per rifornire di energia solo 50 milioni di veicoli dal rendimento e dalle prestazioni uguali a quelli dei veicoli attualmente in circolazione.

Per completare il quadro, bisogna dire ancora qualcosa sulla pila a combustibile. È vero che le pile a combustibile possono operare con un rendimento netto del 40-50 per cento, e che potrebbero erogare direttamente la loro energia a un motore e a un combinatore per un rendimento termico globale di almeno il 25 e forse anche del 40 per cento. Ma, a parte il costo e la durata, tutte le tendenze attuali indicano che gli apparati a pila a combustibile sono così pesanti da dover essere considerati fuori gara per i veicoli futuri.

Riepiloghiamo ora brevemente. Come si è visto, il venerando motore a ciclo Otto è difficile da battere, se non per quel che riguarda le emissioni di scarico. Per di più questo motore è stato «pulito» in maniera notevole, anche se (con una o due eccezioni) il compito ha richiesto congegni catalitici che tendono a deteriorarsi con l'uso. Per il resto il motore convenzionale è leggero, ha un rendimento accettabile ed è estremamente economico da costruire.

Un'alternativa promettente ai catalizzatori per ridurre i gas di scarico del motore a ciclo Otto è una modificazione della camera di combustione e del sistema di

	RENDIMENTO DEL MOTORE (PERCENTUALE) PIENO CARICO	25 PER CENTO DEL CARICO	10 PER CENTO DEL CARICO	RAPPORTO POTENZA-MASSA CHILOWATT PER CHILOGRAMMO	SCARICHI	COSTO DI FABBRICAZIONE
MOTORE OTTO	30 / 26	27 / 18	24 / 15	0,8 / 0,3	ACCETTABILI CON REGOLATORI; PROBABILITÀ DI DEGENERARE	MOLTO BASSO
MOTORE DIESEL	36 / 26	35 / 20	32 / 18	0,5 / 0,2	ALTO TENORE DI OSSIDI DI AZOTO (NO_x)	MODERATO
MOTORE RANKINE	30 / 20	26 / 18	22 / 15	0,5 / 0,2	MOLTO BUONE; IMPROBABILITÀ DI DEGENERARE	ALTO
MOTORE STIRLING	42 / 30	40 / 30	38 / 28	0,5 / 0,2	MOLTO BUONE; IMPROBABILITÀ DI DEGENERARE	ALTO
MOTORE BRAYTON A CICLO APERTO (TURBINA A GAS)	44 / 25	30 / 10	25 / 8	1 / 0,6	MOLTO BUONE; IMPROBABILITÀ DI DEGENERARE	MODERATO; POTENZIALMENTE BASSO (MATERIALE CERAMICO)
MOTORE BRAYTON A CICLO CHIUSO	36 / 22	36 / 22	34 / 20	1 / 0,2	MOLTO BUONE; IMPROBABILITÀ DI DEGENERARE	MOLTO ALTO

Non è da escludere che i sistemi di propulsione proposti superino, dal punto di vista del rendimento, i motori a ciclo Otto, ma i costi di fabbricazione prospettati in proposito sono alti, con la possibile eccezione della turbina a gas. Per il momento il loro pregio maggiore (eccezion fatta per il Diesel) è costituito dalla minore quantità di gas di scarico. Il rendimento viene indicato su una scala di valori, in cui la cifra più bassa rappresenta il rendimento possibile con la tecnologia attuale e la cifra più alta rappresenta quello che forse si potrebbe ottenere fra dieci anni nei motori più avanzati; lo stesso metodo è stato usato per il rapporto fra la potenza e la massa. Tutte le cifre relative al motore Brayton a ciclo chiuso sono solo speculative, poiché il motore in questione non è mai stato costruito in una forma adatta per gli autoveicoli.

Un motore Stirling sperimentale è stato sviluppato congiuntamente dalla Ford Motor Company negli Stati Uniti e dalla N. V. Philips Gloeilampenfabrieken in Olanda. Poiché il motore Stirling brucia ininterrottamente il combustibile in condizioni ottimali, caratteristica, questa, che esso condivide con i motori a vapore e con le turbine a gas, il suo scarico ha un basso tenore di sostanze inquinanti. Potenzialmente il motore Stirling ha, a carico parziale, un rendimento più alto di quello di qualsiasi concorrente, un fattore importante per i veicoli che percorrono le autostrade.

iniezione del carburante che consente un processo di combustione a due fasi. Una miscela molto ricca di carburante e aria viene prima accesa e poi utilizzata per bruciare una miscela molto povera. In questo modo viene ad abbassarsi la temperatura massima della combustione, con una conseguente riduzione della produzione di NO_x. I motori di questo tipo sono noti col nome di motori a carica stratificata. Uno, costruito dalla Honda, l'Honda CVCC, è in produzione su scala commerciale dal 1974, e altri sono tuttora in fase di sviluppo.

Se tali motori potranno avere o meno la meglio sui motori più convenzionali, modificati in modo da bruciare miscele uniformi molto povere, è qualcosa che non si è ancora in grado di prevedere in maniera attendibile.

L'alternativa principale al motore a ciclo Otto è il motore con accensione a compressione, il Diesel, disponibile da molti anni sulle Mercedes e sulle Peugeot, e ora offerto anche sulle Volkswagen, sulle Oldsmobile, sulle Cadillac, sulle Opel, le Ford e anche in Italia su alcune Fiat e Alfa Romeo. I motori Diesel attuali sono meno allettanti di quelli a ciclo Otto per tre motivi principali. Primo, sono più pesanti a parità di potenza; secondo, sono più costosi (vengono a costare circa il doppio per cavallo vapore); e, terzo, non sembrano in grado di rispettare la severa regolamentazione sui livelli di NO_x, che entrerà in vigore per quanto riguarda gli Stati Uniti nel 1980. Altri inconvenienti minori sono costituiti da una messa in moto difficile alle basse temperature, da una gamma più ridotta di velocità operative, da una maggiore rumorosità e dall'odore sgradevole dello scarico, con la possibilità, che incombe come una minaccia, di un certo grado di tossicità.

Ciò nonostante, il motore Diesel può essere migliorato. Un metodo, usato in molti autocarri pesanti e introdotto nei «pulmini» della General Motors, è quello di accoppiare il motore Diesel a una piccola turbina, la quale sfrutta il gas di scarico del Diesel per azionare un compressore che aumenta la pressione dell'aria diretta alla camera di combustione del motore. Una disposizione del genere abbassa il rapporto di compressione richiesto dal Diesel, riduce il peso globale del motore a parità di lavoro utile, e probabilmente consente anche una riduzione del costo globale.

Un'altra proposta è quella di riprogettare il Diesel automobilistico per un ciclo a due tempi, cosa, anche questa, già in uso su molti autocarri e autobus, in modo da iniettare il carburante a ogni giro dell'albero a gomiti invece che ogni due giri, come invece avviene nello schema di tipo convenzionale. Per i motori piccoli però il vantaggio di questo sistema sembra dubbio. Sono allo studio anche dispositivi per far aumentare la turbolenza dell'aria nella camera di combustione nel momento in cui viene iniettato il carburante, allo scopo di ridurre le emissioni e il cattivo odore dello scarico, nonché il battito un po' irritante del Diesel e la pressione massima necessaria per l'accensione. Il risultato netto potrebbe essere una riduzione del peso del motore, pagando un piccolo scotto al rendimento. Quello che rimane da vedere è se questi o altri accorgimenti tecnici riusciranno a trasformare il Diesel in un motore d'automobile pienamente competitivo.

Anche se nessun produttore di primo piano sta lavorando in questa direzione, il motore a vapore ha sempre i suoi devoti sostenitori. I motori attualmente in fase di sviluppo lavorano con il vapor d'acqua, piuttosto che con vapori organici, perché, potendo raggiungere temperature più elevate, si ottiene un grado più alto di rendimento e si possono ridurre le dimensioni del condensatore. Il rendimento peraltro rimane basso, e l'unica vera speranza di un netto miglioramento sta nell'impiegare materiali piuttosto costosi che rendano possibile aumentare la temperatura del vapore. C'è infine il fastidioso problema del congelamento dell'acqua quando la temperatura ambiente è molto bassa, un problema per il quale non si è riusciti ancora a trovare una soluzione del tutto soddisfacente.

Il motore Stirling è in fase di sviluppo congiuntamente presso la Ford Motor Company negli Stati Uniti e presso la N. V. Philips Gloeilampenfabrieken in Olanda. Qui il fluido operativo preferito è l'idrogeno, perché esso trasferisce il calore meglio di qualunque altro gas e offre il minimo di attrito. Il rendimento a carico parziale di alcuni motori Stirling sperimentali è più elevato (40 per cento) di quello di qualsiasi concorrente attuale. Il motore ha l'ulteriore vantaggio, condiviso dalle turbine a gas e a vapore, di essere in grado di operare con quasi ogni combustibile. È molto silenzioso, è il più facile da mettere in moto di tutti i motori proposti in alternativa al motore a ciclo Otto e non ha problemi di congelamento. Siccome gli attuali modelli sperimentali hanno bisogno di una notevole quantità di metallo capace di resistere alle alte temperature, i costi di produzione sarebbero elevati. C'è comunque la speranza che si possano costruire componenti alternativi di materiale ceramico. Avendo un ciclo a gas, il motore Stirling è in grado, potenzialmente, di operare a temperature e a pressioni ancora più elevate di quelle a cui opera attualmente, il che lo renderebbe ancora più leggero e ne aumenterebbe il rendimento.

I motori più leggeri di cui si possa disporre attualmente sono le turbine a gas. Con le temperature operative già in vista, il rendimento termico della turbina a gas può diventare il più elevato fra quelli di tutti i motori progettati in alternativa al motore a ciclo Otto raggiungendo forse valori pari al 55 per cento. D'altra parte, la turbina a gas a ciclo aperto ha probabilmente il più basso grado di rendimento a carico parziale. Il rendimento a carico parziale migliora con l'aggiunta di uno scambiatore di calore e con la diminuzione del rapporto di pressione. Se il ciclo è completamente chiuso, il rendimento a carico parziale è secondo solo a quello del motore Stirling. Queste turbine a ciclo chiuso, che sono state proposte seriamente ma che non sono ancora state costruite in modo da adattarsi al campo automobilistico, avrebbero un rendimento termico minore rispetto alle turbine a ciclo aperto, perché il calore dovrebbe essere trasferito attraverso pareti, quelle del riscaldatore, sottoposte a un'estrema sollecitazione, e perché si renderebbe necessario, all'uscita del ciclo, un dispositivo di raffreddamento, il quale farebbe salire la temperatura di dispersione del calore oltre quella di una turbina a ciclo aperto. La mia opinione personale è che il motore Stirling costituisce per le automobili una prospettiva migliore della turbina a gas a ciclo chiuso.

La turbina a gas caratterizzata da un basso rapporto di pressione e dotata di uno scambiatore di calore (o rigeneratore) è quanto di meglio si possa prospettare in fatto di turbine a gas per autoveicoli, e grandi sforzi sono stati fatti per il suo sviluppo, specie da parte della Chrysler Corporation. Nella sua forma attuale, essa richiede notevoli quantità di materiali resistenti alle alte temperature, e pertanto il suo costo sarebbe molto elevato. Si sta lavorando attivamente per cercare di utilizzare materiali ceramici nelle parti calde del motore, con la possibilità di raggiungere, all'entrata della turbina, temperature dell'ordine di 1650 gradi centigradi. Contemporaneamente verrebbero a diminuire anche i costi, perché i materiali ceramici dovrebbero costare, al chilogrammo, sì e no un ventesimo rispetto alle leghe in grado di resistere al calore. I costi di fabbricazione diminuirebbero in proporzione, perché, quando si arriva alla produzione in serie, i costi di fabbricazione di qualsiasi motore calano a solo un piccolo multiplo del costo dei materiali. Il rendimento a carico parziale di una turbina a gas di questo tipo dovrebbe superare quello del motore convenzionale con accensione a scintilla, per cui la turbina sarebbe difficile da battere sia per l'applicazione in campo automobilistico, sia per molte altre applicazioni.

Fin qui ho esposto la mia tesi, secondo la quale la scelta di un'alternativa al motore con accensione a scintilla, attualmente predominante, è estremamente complessa. Nel caso specifico statunitense (ma queste osservazioni si adattano, con le necessarie modificazioni, alle situazioni di tutti i paesi industrializzati), se il governo dovesse dare il suo appoggio soltanto a uno o due progetti alternativi, la cosa sarebbe non solo scorretta, ma anche, forse, controproducente, a causa dell'effetto inibitorio sulle altre linee di ricerca e progettazione. Quale dovrebbe essere allora il ruolo del governo?

Il governo ha due principali possibilità di scelta, quando si trova di fronte ad attività che comportano quelli che gli economisti chiamano grandi costi sociali, quei costi cioè che devono essere sostenuti non dal singolo consumatore, ma dalla società nel suo insieme. Il prezzo di listino di un'automobile che inquina l'aria che tutti respiriamo non include una valutazione delle molteplici conseguenze di tale inquinamento. Quel costo viene scaricato all'esterno. Analogamente si può sostenere che c'è un costo scaricato all'esterno anche quando gli automobilisti «distruggono» in pratica una parte cospicua di una risorsa non rinnovabile come il petrolio.

In situazioni del genere il governo può scegliere la via della regolamentazione o quella della tassazione. La prima è chiaramente preferibile quando i costi totali di una regolamentazione rimangono entro limiti accettabili, o quando i costi della tassazione toccano livelli inaccettabili. Per questa ragione, se il governo USA avesse posto limiti ragionevoli alle emissioni di scarico e obiettivi facilmente raggiungibili in fatto di economia di carburante, i costi della regolamentazione sarebbero stati accettabili. Di fatto, però, con l'Energy Policy and Consevation Act (legge sulla conservazione dell'energia e sulla politica energetica) i limiti fissati agli scarichi sono quasi proibitivi, e le mete relative all'economia di carburante sono pressoché irraggiungibili. Stando così le cose, i costi dell'altra alternativa, quella cioè della tassazione, sarebbero stati quasi certamente inferiori. Il governo avrebbe potuto fissare un'imposta progressiva sugli scarichi di ogni modello di automobile e un'imposta sull'energia che aumentasse il prezzo del carburante, o forse un'imposta comparabile sul risparmio di energia misurato sul consumo di carburante di ogni singolo modello.

Una politica del genere avrebbe fatto ricadere all'interno, sia pure in modo approssimativo, i costi sociali dell'inquinamento e del consumo dissipatore di una risorsa che va man mano scomparendo. I consumatori per la maggior parte sceglierebbero, compatibilmente con le proprie esigenze di trasporto, le automobili con le imposte più basse. I fabbricanti, dal canto loro, senza più le incertezze del mutare delle norme, svilupperebbero presumibilmente quei motori che presentano maggiori attrattive per i consumatori. Gli incentivi offerti a chi compra e a chi vende tenderebbero a indirizzare lo sforzo dello sviluppo del motore verso un optimum sociale. Per completare questo stretto intrecciarsi di azioni e reazioni, bisognerebbe che il ricavato delle imposte venisse ripartito direttamente fra le persone sulle quali grava oggi il peso dei costi sociali, e queste persone costituiscono in pratica l'intera popolazione. In altre parole, il ricavato delle imposte dovrebbe essere ripartito in maniera paritaria fra tutti i cittadini adulti degli Stati Uniti.

I problemi logistici della produzione automobilistica rendono impossibile l'introduzione su vasta scala di qualsiasi motore in alternativa al motore a ciclo Otto prima degli anni novanta. A mio parere, la maniera ideale di scegliere uno o più motori futuri si basa, più che su una regolamentazione, su un mercato libero modificato, creato dalla tassazione.

PROTEZIONE DELL'AMBIENTE

LE SCIENZE
edizione italiana di
SCIENTIFIC AMERICAN
ha finora pubblicato su questo argomento i seguenti articoli:

LA SICUREZZA DEGLI IMPIANTI NUCLEARI
di P. Giuliani e D. Maniori (n. 15)

CIRCOLAZIONE GLOBALE DELL'INQUINAMENTO ATMOSFERICO
di R.E. Newell (n. 32)

NUOVE TECNOLOGIE PER LO STUDIO DELL'INQUINAMENTO DEL MARE
di C. Morelli (n. 35)

INQUINAMENTO DA MERCURIO
di L.J. Goldwater (n. 36)

I MODELLI MATEMATICI E L'AMBIENTE NATURALE
di R. Pennacchi (n. 45)

ENERGIA « PULITA » DA COMBUSTIBILI « SPORCHI »
di A.M. Squires (n. 53)

L'ELIMINAZIONE DEI RIFIUTI NELL'OCEANO
di W. Bascom (n. 76)

L'INQUINAMENTO DA TRIZIO
di G. Zamparo (n. 90)

EFFETTI ANTROPOGENICI SULL'OZONOSFERA
di F. Verniani (n. 102)

Il differenziamento chimico delle cellule nervose

Perché possa svilupparsi l'intricata rete di cellule che costituisce il sistema nervoso ciascuna cellula deve "scegliere" la sostanza trasmettitrice più appropriata alle connessioni che sta per stabilire

di Paul H. Patterson, David D. Potter ed Edwin J. Furshpan

Il flusso d'informazione attraverso il sistema nervoso è determinato dal particolare tipo di connessioni che i neuroni, o cellule nervose, stabiliscono l'uno con l'altro e con altri tessuti dell'organismo come i muscoli e le ghiandole. È un problema di notevole interesse cercare di stabilire come questa rete così complessa si sia formata durante lo sviluppo dell'organismo. Nell'embrione, i neuroni inizialmente separati entrano in contatto tra loro, in parte migrando e in parte emettendo dal corpo cellulare sottili prolungamenti, chiamati assoni e dendriti. Queste fibre di solito si ramificano ripetutamente e possono estendersi verso parti distanti del sistema nervoso oppure lasciare quest'ultimo per innervare i tessuti effettori. Il sito di comunicazione tra due neuroni o tra un neurone e una cellula effettrice è la sinapsi, in cui la terminazione dell'assone del neurone innervante giunge molto vicino alla cellula bersaglio, ma non la tocca: l'intervallo tra le due è ripieno di un liquido e si chiama fessura sinaptica. Con l'arrivo di un impulso nervoso, le molecole di un neurotrasmettitore chimico vengono liberate dalla terminazione nervosa, si spostano attraverso la fessura sinaptica e si combinano con proteine recettrici estremamente specifiche, presenti alla superficie della cellula bersaglio. L'attivazione di tali recettori innesca una risposta elettrica o biochimica.

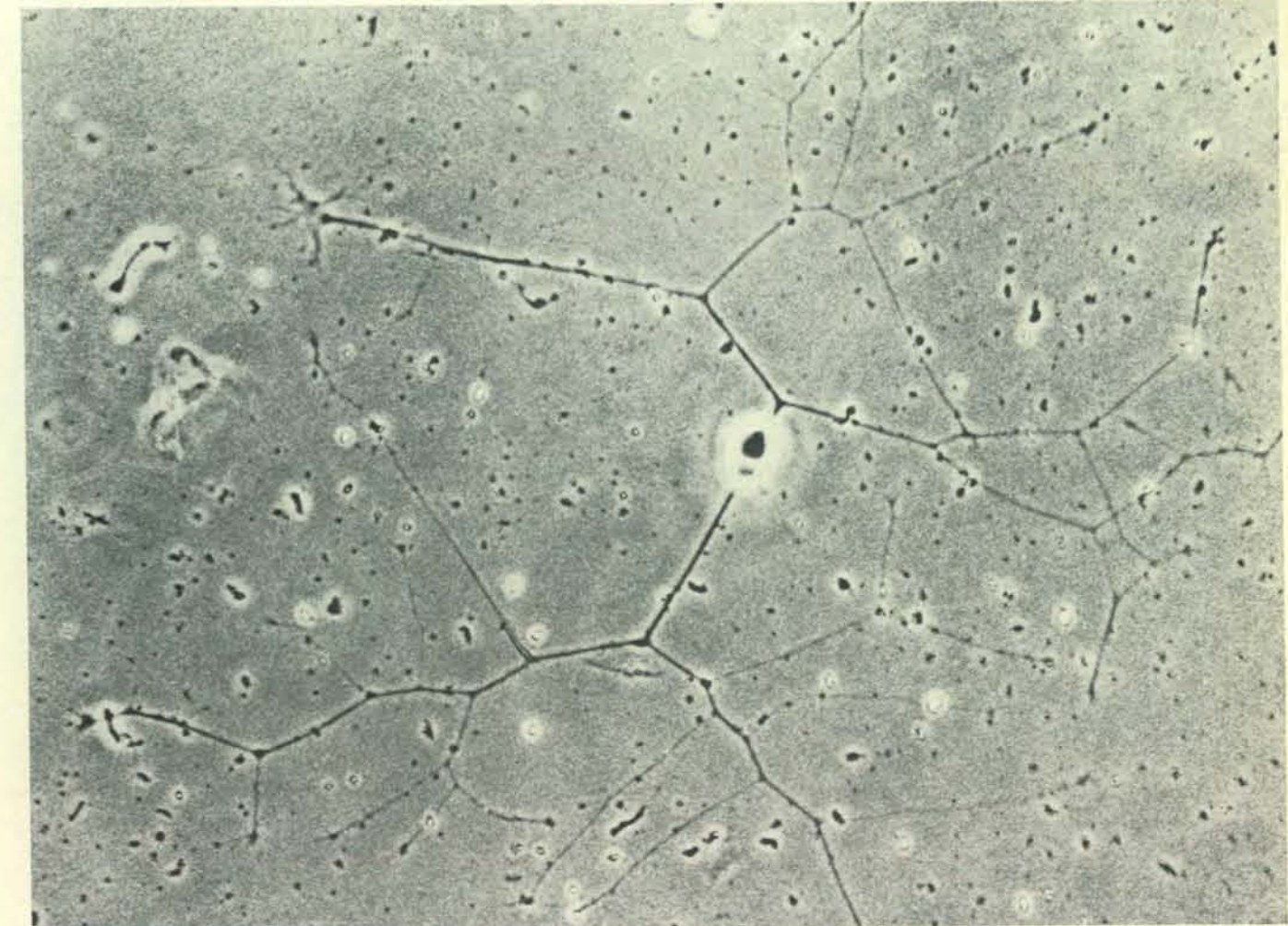
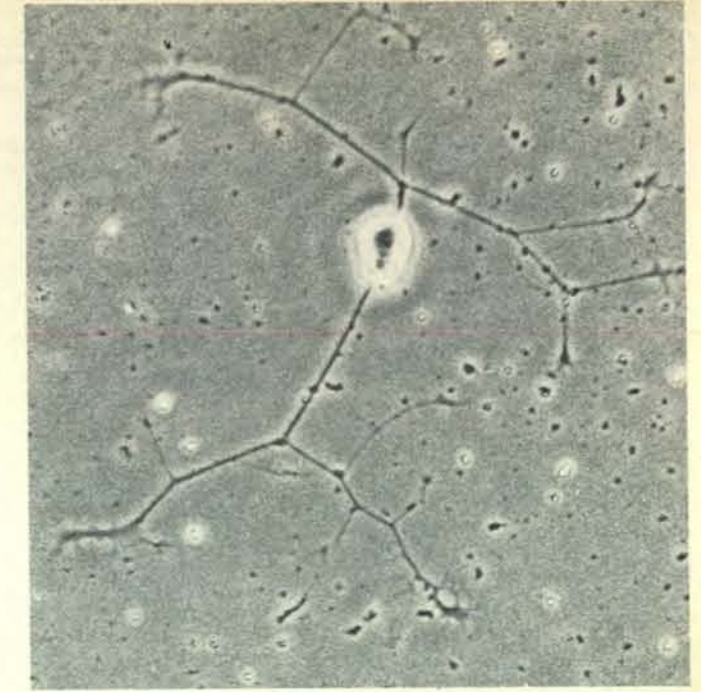
Il fatto che sia la chimica sia la topologia siano implicate nella costruzione del sistema nervoso dà a esso un grado di complessità ben superiore a quello dei circuiti dei calcolatori, costruiti dall'uomo. Si ritiene che più di una dozzina di differenti sostanze possano essere considerate neurotrasmettitori e possano avere sia un effetto eccitatorio (rendendo più attiva la cellula bersaglio) sia un effetto inibitorio (impedendo alla cellula bersaglio di diventare attiva). In che modo i neuroni del sistema nervoso in via di sviluppo «decidono» quale neurotrasmettitore secernere in base alla loro localizzazione e funzione? E come si forma la sinapsi perché vi sia un appaiamento appropriato tra il neurotrasmettitore secreto dal neurone innervante e i recettori presenti sulla cellula bersaglio? L'indagine su questi interrogativi ha richiesto la collaborazione di elettrofisiologi, biochimici e microscopisti elettronici, operanti in parecchi laboratori.

Nel nostro laboratorio alla Harvard Medical School, abbiamo affrontato con 14 altri colleghi il problema generale di come si differenzino chimicamente i neuroni, esaminando lo sviluppo del sistema nervoso autonomo, che regola le attività del sistema circolatorio e di una varietà di organi e di ghiandole. Il vantaggio di scegliere un simile sistema per una tale ricerca è che le sue porzioni periferiche sono facilmente accessibili al ricercatore. Inoltre, l'organizzazione chimica che esso possiede è relativamente semplice: la maggior parte delle cellule nervose che lo costituiscono secernono o acetilcolina o noradrenalina e sono definite rispettivamente colinergiche o adrenergiche. Sembra giustificato ritenere che i meccanismi fondamentali implicati nella «decisione» di un neurone di diventare o adrenergico o colinergico sono analoghi a quelli che determinano quale, tra parecchi possibili neurotrasmettitori, sarà secreto da un neurone nell'encefalo.

Nel sistema nervoso autonomo esistono due tipi principali di vie: quelle simpatiche e quelle parasimpatiche. Questi due sottosistemi hanno spesso azioni antagoniste sullo stesso organo. Per esempio, le cellule muscolari cardiache, che si contraggono spontaneamente e ritmicamente, sono innervate sia da nervi simpatici sia da nervi parasimpatici. In situazioni di stress, i primi secernono noradrenalina, che fa battere più rapidamente e più energicamente il cuore. In periodi di calma o di inattività, i nervi parasimpatici secernono acetilcolina, che lo fa battere più lentamente.

Ambedue le vie autonome che innervano il cuore hanno inizio nell'encefalo o nel midollo spinale con i neuroni specifici. Questi neuroni centrali non raggiungono direttamente il muscolo cardiaco, ma producono sinapsi colinergiche eccitatorie su neuroni posti in un ganglio che rappresenta la stazione nervosa esterna al midollo spinale: sono questi ultimi neuroni che, a loro volta, innervano il muscolo cardiaco. Quindi la via simpatica è formata da un neurone centrale colinergico e da un neurone gangliare adrenergico, ambedue eccitatori; la via parasimpatica è formata, invece, da un neurone centrale e da un neurone gangliare colinergici, il primo eccitatorio e il secondo inibitorio. (La capacità dell'acetilcolina di eccitare in corrispondenza della prima sinapsi e di inibire in corrispondenza della seconda è dovuta a differenze nella composizione chimica delle cellule bersaglio.)

Che cosa fa sì che un determinato neurone in una via del sistema nervoso autonomo diventi colinergico o adrenergico? Per indagare su questo problema, abbiamo effettuato la maggior parte degli esperimenti su neuroni immaturi, ottenuti da gangli simpatici di ratti neonati e coltivati in vitro. Questo tipo di preparato è stato scelto per tre ragioni. Innanzitutto, il ganglio simpatico adulto contiene sia neuroni adrenergici sia un numero più piccolo di neuroni colinergici (che innervano certi vasi sanguigni e le ghiandole sudoripare), per cui nello stesso sistema si può studiare il differenziamento dei due tipi di neurone. In secondo luogo, come è stato dimostrato da Rita Levi-Montalcini e dai suoi colleghi della Washington University, lo sviluppo normale dei neuroni simpatici nell'animale integro dipende dalla presenza di una particolare proteina, chiamata fattore di crescita del nervo (NGF): quando questo NGF viene aggiunto al terreno colturale, si riesce a coltivare in vitro tali neuroni.

Dennis Bray, che oggi lavora presso la Cell Biophysics Unit del Medical Re-

Un neurone in coltura ottenuto da un ganglio simpatico di ratto neonato emette sottili prolungamenti ramificati quando viene incubato in presenza di una proteina specifica, chiamata fattore di crescita del nervo. La microfotografia in alto a sinistra mostra la cellula dopo 240 minuti di incubazione; quella in alto a destra dopo 300 minuti e quella in basso dopo 500 minuti. I prolungamenti crescono ininterrottamente e, alla fine, formano una rete sulla superficie della piastra di coltura. Possono anche formare delle sinapsi con altri neuroni o con cellule muscolari in coltura. Il vantaggio offerto dallo studio dei neuroni in vitro consiste nel fatto che l'ambiente può essere controllato con cura. Le microfotografie qui riprodotte sono state scattate da Dennis Bray della Cell Biophysics Unit del Medical Research Council di Londra.

search Council di Londra, ha successivamente messo a punto condizioni di coltura specifiche, in cui sopravvivevano i neuroni simpatici del ratto neonato, ma non le cellule di altro tipo del ganglio: in questo caso i suddetti neuroni si sviluppavano nella virtuale assenza di qualsiasi altro tipo di cellula. Inoltre, quando vengono rimossi dai gangli simpatici del ratto neonato, i neuroni sono ancora immaturi per molti aspetti fondamentali: la maggior parte di essi non ha ancora stabilito connessioni sinaptiche e si trova in uno stadio precoce di differenziamento del metabolismo del neurotrasmettitore.

Il vantaggio di studiare lo sviluppo dei neuroni in coltura piuttosto che nell'animale in toto sta nel fatto che il ricercatore ha un miglior controllo sull'ambiente cellulare e sul liquido che circonda i neuroni e può quindi indagare con maggiore facilità sui fattori chiave per l'induzione o per il controllo di particolari funzioni cellulari. Naturalmente, i neuroni in coltura devono essere riforniti di quelle sostanze nutritizie che normalmente riceverebbero dal sangue, per esempio amminoacidi, zuccheri e vitamine. Un buon equilibrio tra ossigeno e anidride carbonica, che di solito viene mantenuto dai polmoni, si realizza grazie a uno scambio di gas tra il terreno di coltura e l'ambiente controllato dell'incubatrice dove viene posta la coltura. La rimozione delle scorie tossiche, in genere effettuata dai reni, viene effettuata sostituendo periodicamente il terreno di coltura. Dato che le cellule isolate non hanno alcuna difesa immunologica contro i microrganismi, è necessario che strumenti, piastre di coltura e terreni di coltura siano sterili. Infine, per facilitare l'aderenza degli assoni e dei dendriti al substrato, la plastica di cui è fatta la superficie della piastra di coltura viene rivestita di un sottile strato di collagene, costituente naturale della matrice intercellulare.

I neuroni simpatici sopravvivono in presenza del fattore di crescita del nervo, ma non in sua assenza, ed emettono prolungamenti che continuano ad allungarsi per molte settimane. Alla fine, la superficie della piastra risulta ricoperta da una fitta rete di assoni e di dendriti e, fatto sorprendente, i neuroni sviluppatisi in questo ambiente esclusivamente artificiale, in assenza di cellule non neuroniche, acquisiscono molte caratteristiche dei neuroni simpatici normali e lo fanno in un periodo di tempo simile a quello che si osserva nell'animale integro. I neuroni coltivati generano impulsi, possiedono recettori di superficie per l'acetilcolina (il trasmettitore liberato dai neuroni del sistema nervoso centrale, che di norma li innervano) e sviluppano la capacità di sintetizzare la noradrenalina dal suo precursore, l'amminoacido tirosina. Di fatto, la maggior parte, se non tutti i neuroni dei gangli simpatici del ratto sintetizzano la noradrenalina prima della nascita. Questi risultati fanno pensare che, prima di essere rimossi dall'animale, i neuroni abbiano ricevuto dei segnali di sviluppo, delle istruzioni su come utilizzare la noradrenalina come trasmettitore. In presenza di NGF, che permette alle cellule di sopravvivere in coltura, leggono semplicemente il loro specifico programma genetico.

In assenza di cellule non neuroniche, i neuroni simpatici in coltura sono anche in grado di formare sinapsi l'uno con l'altro. Come è stato illustrato da Richard P. Bunge e collaboratori alla Washington University School of Medicine, queste sinapsi assomigliano come struttura a quelle che stabiliscono i neuroni adrenergici adulti, nel senso che la noradrenalina viene accumulata in minuscole vescicole, o sacchetti, le quali presentano una porzione centrale opaca, quando sono colorate con permanganato per essere osservate al microscopio elettronico. Tuttavia, pur essendo la noradrenalina liberata e assunta in corrispondenza di queste sinapsi, essa, come neurotrasmettitore, non ha alcun effetto riconoscibile sull'attività elettrica del neurone bersaglio, il che significa che o i neuroni in coltura non possiedono recettori per questa sostanza, oppure, se tali recettori sono presenti, che non riescono a evocare un segnale elettrico significativo. Pochi sono i casi noti di sinapsi silenziose dal punto di vista elettrico, come queste.

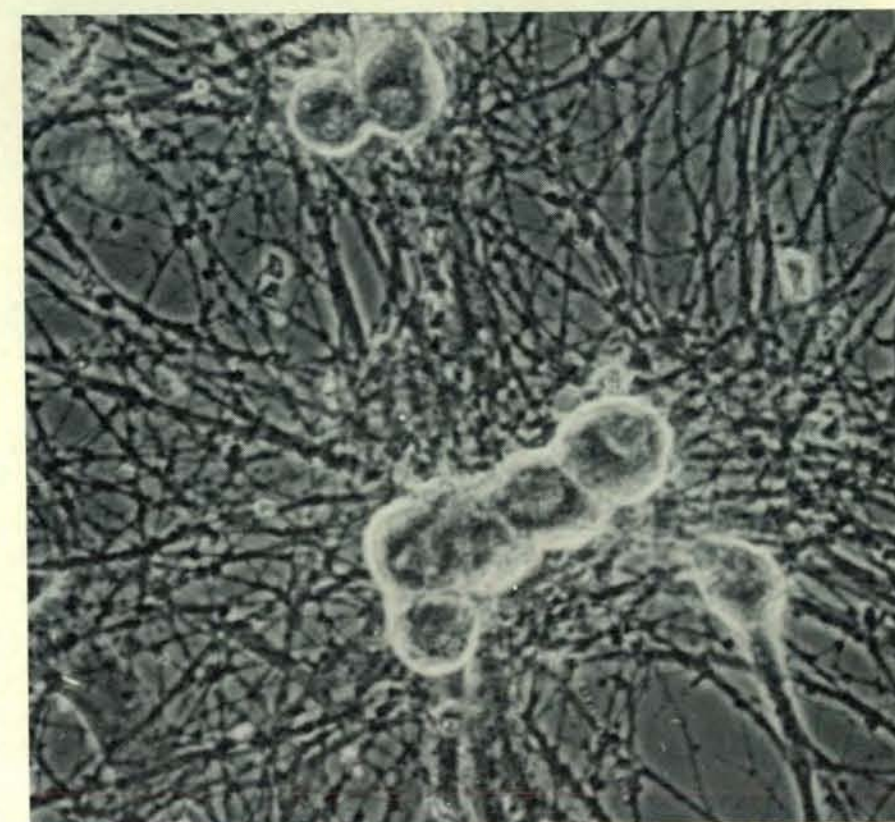
Quando i neuroni simpatici vengono coltivati con cellule di origine gangliare diverse dai neuroni, si osservano sorprendenti differenze nel loro differenziamento chimico. Invece di esprimere soltanto funzioni adrenergiche, le colture miste sintetizzano e accumulano elevate concentrazioni di acetilcolina, indicando così che molti neuroni sono diventati colinergici. L'indagine elettrofisiologica ha rivelato che un'importante frazione di neuroni forma anche le sinapsi colinergiche con i neuroni vicini, sostituendosi così al normale *input* colinergico proveniente dai neuroni del sistema centrale, che sono assenti dalla coltura. Diversamente dalle sinapsi adrenergiche, formate in assenza di cellule non neuroniche e silenziose dal punto di vista elettrico, queste sinapsi colinergiche sono attive elettricamente e hanno un'azione eccitatoria.

L'effetto della presenza di cellule diverse dai neuroni sulla scelta del trasmettitore è esercitato mediante contatto diretto con i neuroni simpatici oppure grazie alla liberazione di qualche fattore chimico nel terreno di coltura? Per rispondere a questo interrogativo abbiamo coltivato i neuroni su una serie di piastre e le cellule di tipo diverso su un'altra serie. Ogni due giorni, il terreno che era stato «condizionato», facendo incubare in esso il secondo gruppo di cellule, veniva trasferito alle colture di neuroni. Abbiamo riscontrato che il trattamento con questo terreno bastava a indurre le proprietà colinergiche nei neuroni simpatici: le cellule sintetizzavano e accumulavano acetilcolina e la secernevano in corrispondenza delle sinapsi colinergiche funziona-

TRONCO CEREBRALE
FIBRA ADRENERGICA
FIBRA COLINERGICA
ALTRE FIBRE
SISTEMA PARASIMPATICO
NEURONE CENTRALE
GANGLIO CARDIACO
NEURONE GANGLIARE
ALLA GHIANDOLA SUDORIPARA
NEURONE SPINALE
NEURONE GANGLIARE
SISTEMA SIMPATICO
MIDOLLO SPINALE
GANGLI SIMPATICI

a
ASSONE COLINERGICO
VESCICOLE SINAPTICHE CHIARE
RECETTORI
MOLECOLE DI ACETILCOLINA
MUSCOLO CARDIACO

b
ASSONE ADRENERGICO
VESCICOLE SINAPTICHE DENSE
MOLECOLE DI NORADRENALINA
RECETTORI
MUSCOLO CARDIACO

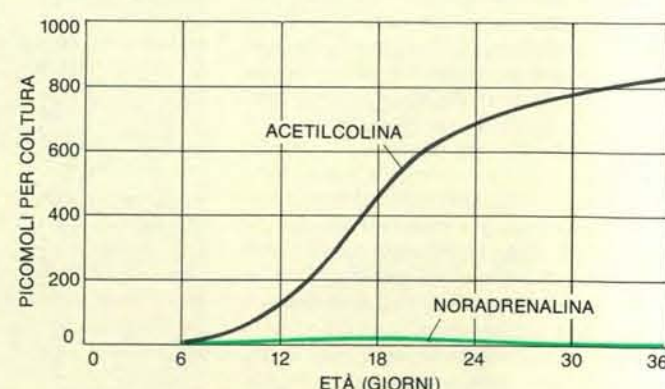
Due vie autonome con azioni antagoniste controllano la frequenza del battito cardiaco. Quando l'organismo è sotto stress e il cuore batte più in fretta, è in attività il sistema simpatico. Le vie riflesse eccitano i neuroni simpatici nel midollo spinale; questi inviano dei prolungamenti ai neuroni presenti in un ganglio, una stazione nervosa esterna al midollo spinale. Quando questi neuroni gangliari vengono eccitati dall'input spinale, fanno battere più in fretta il cuore mediante secrezione di noradrenalina sulle cellule del muscolo cardiaco. Per contro, quando il corpo si rilassa e il cuore batte più lentamente, è in attività il sistema parasimpatico: le vie riflesse eccitano i neuroni parasimpatici presenti nel midollo allungato e questi inviano prolungamenti a neuroni bersaglio, presenti nel ganglio cardiaco del cuore. I neuroni gangliari, a loro volta, fanno battere più lentamente il cuore grazie alla secrezione di acetilcolina sulle cellule del muscolo cardiaco. La secrezione della sostanza neurotrasmettitrice avviene in corrispondenza di varicosità, piccoli rigonfiamenti disposti a intervalli lungo gli assoni che secernono la noradrenalina (assoni adrenergici) e l'acetilcolina (assoni colinergici), come si può vedere schematizzato nei riquadri all'estrema destra.

I neuroni immaturi, ottenuti da gangli simpatici di un ratto neonato, si possono coltivare su piastra, come risulta da queste microfotografie di colture, scattate con un microscopio in contrasto di fase. Nel mezzo standard, le cellule di sostegno del ganglio sopravvivono e si dividono, formando alla fine uno strato continuo intorno e sotto i neuroni *(in alto)*. Le colture che contengono solo neuroni *(in basso)* si possono ottenere facendo crescere cellule gangliari in un mezzo in cui sopravvivono i neuroni, ma non le cellule di altro tipo. Uno di questi mezzi contiene un analogo di un precursore dell'acido nucleico, il citosinarabinoside, che è tossico quando viene incorporato nelle molecole di acido nucleico delle cellule diverse dai neuroni che si dividono, mentre non ha alcuna influenza sui neuroni, i quali non si dividono come le altre cellule, ma semplicemente si ingrossano.

li che esse formavano l'una con l'altra.

L'effetto del terreno condizionato era dipendente dalla dose: quanto maggiore era la quantità aggiunta al terreno di coltura dei neuroni, tanta più acetilcolina le cellule producevano e tanto più era probabile che formassero sinapsi colinergiche l'una con l'altra. Nel contempo, i neuroni producevano meno noradrenalina e meno sinapsi adrenergiche. Così, nella popolazione di neuroni nel suo insieme, l'espressione delle funzioni colinergiche era grosso modo reciproca dell'espressione delle funzioni adrenergiche. Inoltre, il differenziamento del metabolismo colinergico in presenza del terreno di coltura condizionato richiedeva un tempo simile a quello del metabolismo adrenergico in assenza del mezzo condizionato. Abbiamo ottenuto di recente la dimostrazione che l'ingrediente attivo di tale terreno è una grossa molecola e stiamo ora compiendo tentativi per purificarla.

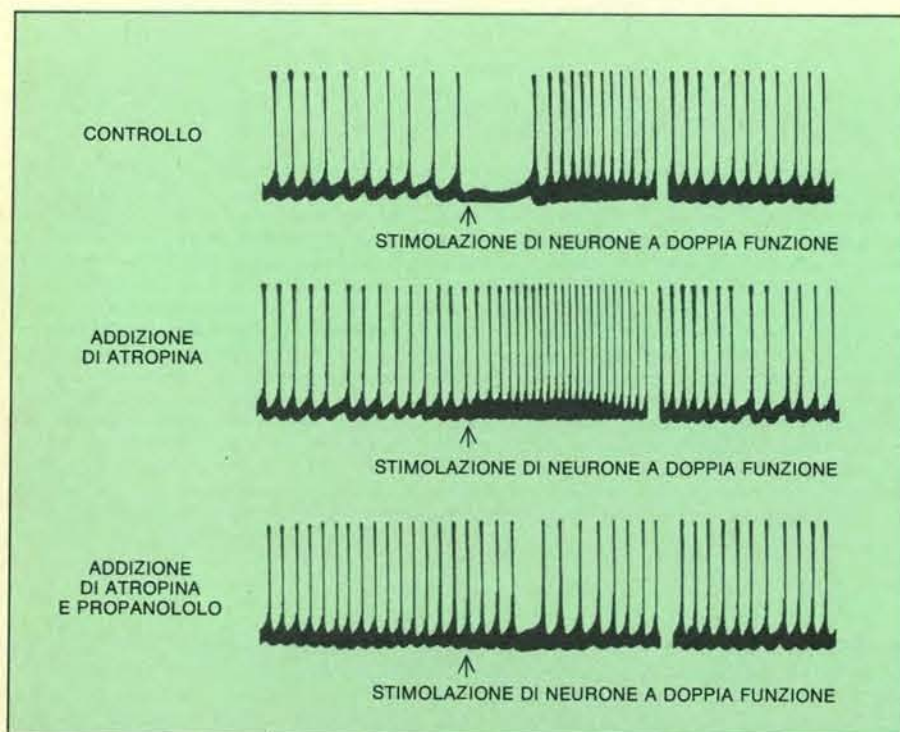
In che modo questo fattore dello sviluppo, prodotto dalle cellule diverse dai neuroni, influisce sulla scelta del neurotrasmettitore? Una possibilità è che, all'inizio, vi siano due popolazioni di neuroni, una destinata a diventare adrenergica e l'altra destinata a diventare colinergica: in questo caso, il fattore condizionante favorirebbe la sopravvivenza e la crescita della popolazione colinergica, mentre diminuirebbe quelle della popolazione adrenergica. Quest'ipotesi è, tuttavia, improbabile, dato che il terreno di coltura condizionato non ha alcun effetto né sulla sopravvivenza né sulla crescita dei neuroni in coltura. Una seconda possibilità è che il terreno condizionato induca l'espressione di proprietà colinergiche in una popolazione colinergica predeterminata, che altrimenti sopravviverebbe, ma non esprimerebbe alcun metabolismo del neurotrasmettitore. Anche quest'ipotesi è stata però scartata perché si è trovato che, in assenza di terreno di coltura condizionato, nessun neurone, virtualmente, era «silenzioso» nei confronti della sintesi del neurotrasmettitore.

Una terza spiegazione per l'effetto del mezzo condizionato è quella che più si conforma alle nostre osservazioni. Secondo tale ipotesi, i neuroni che esprimono proprietà adrenergiche all'inizio sono ancora «plastici» nei riguardi della scelta del neurotrasmettitore per un considerevole periodo di tempo dopo la nascita e possono diventare colinergici sotto l'influsso del terreno di coltura condizionato. Ciò implica che l'ingrediente attivo di tale terreno determina la scelta del neurotrasmettitore e il tipo di sinapsi realizzate da un neurone simpatico, senza influire invece sulla sua sopravvivenza o sull'estensione dei suoi dendriti e dell'assone.

L'effetto dell'ambiente colturale sul differenziamento chimico dei neuroni simpatici è sorprendente. Quando dai gangli simpatici di ratti neonati si estraggono neuroni immaturi e si coltivano in colture pure, quasi tutti sintetizzano e secernono noradrenalina. Tuttavia, se questi neuroni vengono coltivati insieme a cellule che neuroni non sono, oppure se vengono trattati con un terreno colturale che sia stato «condizionato» incubandolo con cellule non neuroniche, una larga maggioranza dei neuroni sitetizzerà e secernerà acetilcolina. I grafici in basso illustrano l'espressione delle funzioni colinergica e adrenergica in presenza e in assenza del mezzo condizionato. I risultati vengono espressi come picomoli (10^{-12} moli) di trasmettitore sintetizzato e accumulato nella coltura di neuroni durante un periodo di incubazione di 4 ore. In assenza del mezzo condizionato *(a sinistra)*, la capacità di sintetizzare noradrenalina insorge dopo un periodo di tre settimane, mentre quella di sintetizzare la acetilcolina rimane trascurabile. Per contro, in colture «sorelle» ottenute con il 62 per cento di terreno condizionato *(a destra)*, la capacità di sintetizzare noradrenalina aumenta subito, ma poi declina fino a un livello basso, mentre aumenta quella di sintetizzare l'acetilcolina. Così, in presenza di elevate concentrazioni di terreno condizionato, l'espressione dei due neurotrasmettitori è reciproca.

Un singolo neurone *(al centro)* estende le proprie ramificazioni su uno strato di cellule cardiache in un sistema microcolturale ideato dagli autori di questo articolo e dai loro colleghi alla Harvard Medical School. Il preparato serve per studiare il differenziamento dei neuroni simpatici nei riguardi della scelta del neurotrasmettitore: questo, secreto dal neurone, può essere identificato grazie al suo effetto sulle cellule del muscolo cardiaco, le quali si contraggono spontaneamente e ritmicamente in coltura. La prima fase nella preparazione della microcoltura consiste nell'ottenere una sospensione di cellule cardiache da ratto neonato. Queste cellule vengono poste su piastra su minuscoli dischi di collageno del diametro di circa 0,5 millimetri, ai quali si attaccano in maniera preferenziale. (Uno di questi dischi compare qui.) Dopo uno o due giorni, quando le cellule cardiache si sono appiattite sul disco, così da formare uno strato continuo, i neuroni vengono anch'essi posti su piastra con una bassa densità, in modo che alcuni dischi ricevano soltanto un neurone. Non appena il neurone ha aderito alla piastra, emette dei prolungamenti filiformi che formano connessioni sinaptiche con alcune cellule del muscolo cardiaco. Il neurotrasmettitore secreto in corrispondenza di queste sinapsi è o acetilcolina, che rallenta il battito delle cellule muscolari cardiache, o noradrenalina che l'accelera. Alcuni neuroni secernono acetilcolina, altri noradrenalina e, a un certo stadio di differenziamento, altri ancora ambedue le sostanze.

La reciprocità nell'espressione delle funzioni adrenergiche e colinergiche può essere ascritta a un certo tipo di «competizione» che avviene all'interno di ogni neurone tra l'istruzione ricevuta prima della nascita che lo spinge a diventare adrenergico e la nuova istruzione che tende a trasformarlo in colinergico. Se esiste davvero una simile competizione, ci si può chiedere se un singolo neurone possa esprimere simultaneamente ambedue i trasmettitori, perlomeno per un breve periodo di tempo. Abbiamo cercato di risolvere l'interrogativo mantenendo in coltura singoli neuroni su uno strato poco esteso di cellule muscolari cardiache per circa un paio di settimane. In simili microculture, il neurone e una cellula cardiaca adiacente sono stati muniti di microelettrodi per segnalare la loro attività elettrica: in questo modo, la scelta del neurotrasmettitore del neurone poteva essere individuata scatenando la liberazione di tale sostanza e osservandone l'effetto sulle cellule cardiache, che pulsano spontaneamente e ritmicamente in coltura. Diversamente dai neuroni simpatici in coltura, le cellule cardiache possiedono sia recettori colinergici sia recettori adrenergici elettricamente attivi: così, un rallentamento o un arresto del battito indica la secrezione di acetilcolina, mentre una maggior frequenza del battito indica la secrezione di noradrenalina. Altre prove sull'identità del trasmettitore si possono ricavare osservando gli effetti di determinati farmaci che competono in maniera specifica con il trasmettitore naturale nel legarsi ai recettori presenti sulle cellule cardiache. Per esempio, l'atropina blocca i recettori cardiaci per l'acetilcolina, mentre il propanololo blocca i recettori cardiaci per la noradrenalina.

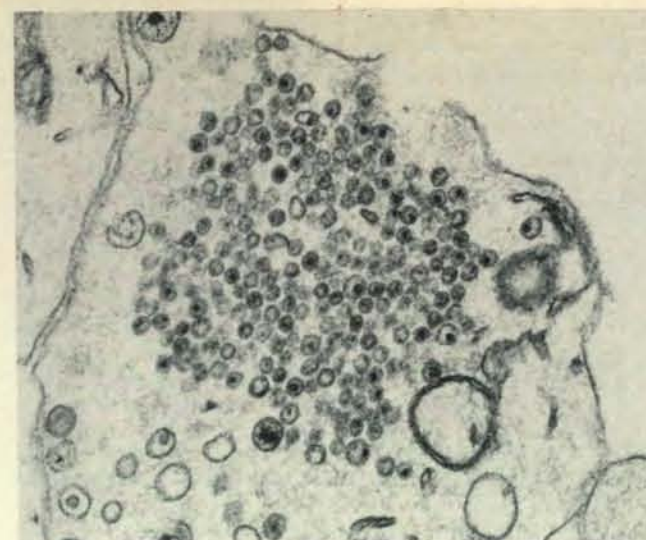
Servendosi di questa tecnica, abbiamo identificato tre tipi di neuroni nelle microcolture di due settimane. Il primo tipo è adrenergico: eccita le cellule cardiache e tale effetto viene bloccato dal propanololo. Il secondo tipo è colinergico: inibisce le cellule cardiache e il suo effetto è bloccato dall'atropina. Il terzo tipo di neuroni, infine, esibisce un'attività sia adrenergica sia colinergica: la sua stimolazione inibisce il cuore con un meccanismo sensibile all'atropina e successivamente accelera il cuore con un meccanismo sensibile al propanololo. Poiché in ogni microcoltura è presente un unico neurone, è chiaro che la stessa cellula è in grado di mediare ambedue gli effetti. Quando le microcolture vengono esaminate al microscopio elettronico, si vedono nei neuroni adrenergici numerose vescicole sinaptiche con la parte centrale più densa, mentre nei neuroni colinergici si notano vescicole chiare e nelle cellule con duplice funzione si notano alcune vescicole con parte centrale densa e molte vescicole chia-

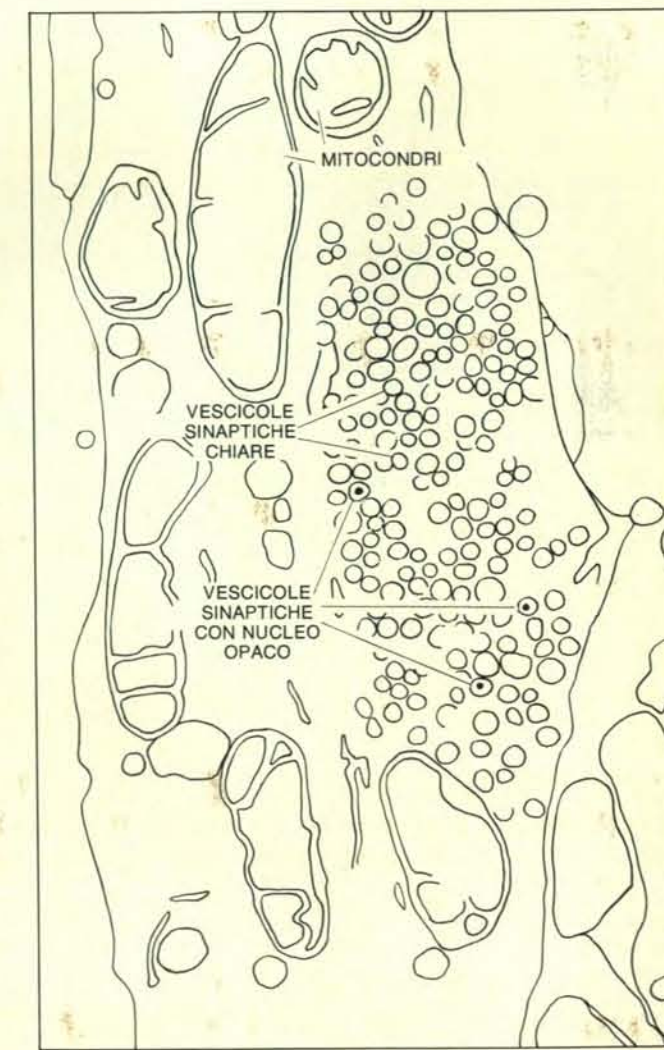
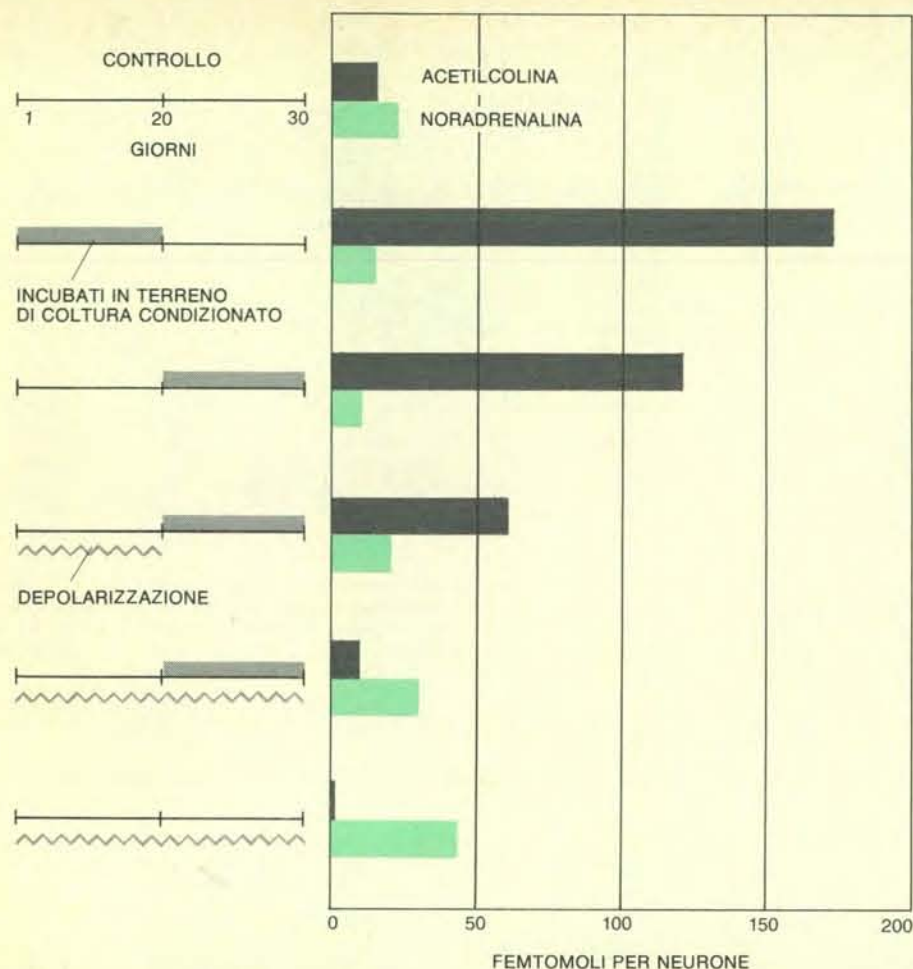
La frequenza di contrazione delle cellule del muscolo cardiaco in una microcoltura contenente un unico neurone rivela il tipo di neurotrasmettitore secreto da quel neurone. In un esperimento tipico, un microelettrodo inserito nel neurone stimola la liberazione del trasmettitore quando un secondo microelettrodo, inserito in una cellula muscolare innervata, segnala le contrazioni di quest'ultima. Il tracciato in alto mostra l'effetto della stimolazione di un neurone colinergico: le contrazioni del muscolo cardiaco cessano temporaneamente. Con la stimolazione di un neurone adrenergico *(in basso)*, la frequenza di contrazione della cellula muscolare innervata aumenta.

Un neurone dalla duplice funzione, presente in una microcoltura di 13 giorni, mostra una combinazione di proprietà colinergiche e adrenergiche. Nel tracciato in alto, la stimolazione del neurone provoca all'inizio una cessazione del battito delle cellule del muscolo cardiaco, cessazione mediata dalla acetilcolina. Questo fenomeno è seguito da una ripresa più vigorosa del battito mediata dalla noradrenalina. Sembra che entrambi i trasmettitori siano secreti contemporaneamente. La separazione temporale dell'inibizione e dell'eccitazione può essere dovuta al fatto che gli effetti della noradrenalina sulle cellule cardiache procedono più lentamente e persistono più a lungo di quelli dell'acetilcolina. Nei due tracciati inferiori, per verificare la duplice funzione, sono stati aggiunti al mezzo colturale farmaci che bloccano in maniera specifica o la trasmissione colinergica o quella adrenergica. Il tracciato al centro mostra l'effetto dell'atropina, che compete con l'acetilcolina per lo stesso recettore: l'inibizione viene rimossa mentre l'eccitazione rimane intatta. Il tracciato in basso mostra l'effetto che consegue a un'aggiunta di propanololo (che compete con la noradrenalina nei confronti dello stesso recettore) alla soluzione di atropina: ambedue gli effetti del neurone, quello eccitatorio e quello inibitorio, ne risultano bloccati.

re. Questi risultati stabiliscono che un solo neurone è in grado di esprimere simultaneamente, in uno stadio immaturo, ambedue i trasmettitori; quello della duplice funzione può sembrare un concetto nuovo, ma di fatto i neuroni in via di sviluppo non erano mai stati analizzati prima in modi che potessero rivelare questo comportamento. In retrospettiva, la duplice funzione è una tappa intermedia logica nel processo di conversione di un neurone adrenergico in uno colinergico sotto l'influsso del terreno di coltura condizionato. Perfino nel caso che non vi sia sovrapposizione temporale tra la sintesi degli enzimi e altri componenti interessati nella sintesi e nella liberazione dei due neurotrasmettitori, sembrerebbe ragionevole ammettere che gli enzimi e le vescicole sinaptiche interessate nella trasmissione adrenergica continuino a funzionare per un certo periodo di tempo dopo che la loro sintesi è cessata. La durata esatta dello stato di duplice funzione non è ancora nota.

Dopo circa quattro settimane, i singoli neuroni in microcoltura sono diventati così grossi che si possono compiere su di essi dei dosaggi biochimici. Le cellule vengono incubate per 8-12 ore in un miscuglio di tirosina radioattiva e di colina radioattiva, quindi vengono determinate le quantità di adrenalina e di acetilcolina che sono state sintetizzate da quelle molecole che fungono da precursori. In assenza di terreno di coltura condizionato, tutti i neuroni producono solo noradrenalina in quantità riconoscibile. In presenza di cellule cardiache, invece, la maggior parte dei neuroni produce soltanto acetilcolina. In nessun caso, dopo quattro settimane in coltura, c'è un numero significativo di neuroni che mostri di avere una duplice funzione o di non averne affatto.

Questi risultati stanno a indicare che, per la maggior parte, i neuroni sono adrenergici quando vengono posti in coltura, ma sono suscettibili a un meccanismo di controllo di tipo *flip-flop*, che determina la loro definitiva scelta del trasmettitore. La durata del periodo di transizione durante il quale si può esprimere la duplice espressione non è nota, ma a quattro o cinque settimane dalla nascita virtualmente tutti i neuroni si sono differenziati in uno stato in cui solo un sistema di trasmettitore si esprime in misura significativa. Nel nostro laboratorio usiamo seguire lo stato dei singoli neuroni per un certo periodo di tempo, misurando a intervalli i loro effetti sulle cellule cardiache. In questo modo, speriamo di riuscire a osservare in maniera diretta la transizione dal comportamento adrenergico al comportamento colinergico e di determinarne anche la durata.

In quale periodo della vita in coltura i neuroni sono più sensibili all'azione del mezzo di coltura condizionato e quanto reversibili sono gli effetti di quest'ultimo? Per trovare una risposta ai suddetti interrogativi, è stato aggiunto terreno condizionato a colture di neuroni simpatici in vari stadi di maturità cellulare a intervalli

Le varicosità sinaptiche lungo gli assoni di neuroni colinergici o adrenergici in coltura mostrano differenze caratteristiche nella loro struttura. Le molecole del neurotrasmettitore vengono immagazzinate in minuscoli sacchi, chiamati vescicole sinaptiche. Le varicosità dei neuroni adrenergici, colorate con permanganato e esaminate al microscopio elettronico, mostrano di contenere vescicole sinaptiche con piccoli nuclei centrali, opachi agli elettroni *(a sinistra)*. Le varicosità dei neuroni colinergici hanno vescicole più chiare al centro *(a destra)*. Le tre microfotografie di questa pagina, che ingrandiscono le strutture 44 000 volte, sono di S. C. Landis della Harvard Medical School.

Una varicosità dalla duplice funzione contiene vescicole con porzione centrale opaca, caratteristiche di un neurone adrenergico *(si veda la mappa)*, con una larga maggioranza di vescicole chiare, tipiche di un neurone colinergico. Così, la struttura della varicosità può essere correlata con i suoi effetti combinati, inibitorio e eccitatorio. Si ritiene che lo stato di duplice funzione sia una fase transitoria durante la conversione di un neurone da adrenergico a colinergico, sotto l'influenza del terreno condizionato. L'esatta durata della duplice funzione non è nota.

L'effetto dell'attività elettrica sul differenziamento chimico dei neuroni simpatici consiste nel ridurre la loro capacità di diventare colinergici quando vengono esposti a un terreno colturale condizionato. Nel ganglio intatto, l'attività risulta dall'innervazione eccitatoria che proviene dai neuroni spinali. Tale influenza eccitatoria può essere imitata in coltura trattando i neuroni simpatici con concentrazioni elevate di ioni potassio, che li depolarizzano. Nell'esperimento qui illustrato, al giorno 0 o al giorno 20 è stato aggiunto potassio o terreno di coltura condizionato e ogni trasmettitore sintetizzato dai neuroni è stato saggiato al giorno 30. (Una femtomole equivale a 10^{-15} moli.) La depolarizzazione delle cellule con potassio ha ampiamente bloccato l'effetto del terreno condizionato. Quindi la minoranza dei neuroni presenti nel ganglio simpatico intatto — che sono colinergici — non possono diventare attivi da un punto di vista elettrico fino a quando non hanno subito un'induzione colinergica da parte di cellule diverse dai neuroni.

di dieci giorni e abbiamo trovato che la reattività delle cellule allo stimolo impartito dal mezzo di coltura condizionato calava rapidamente con il crescere in età delle cellule, raggiungendo un livello molto basso quando lo stimolo veniva somministrato tra il quarantesimo e il cinquantesimo giorno. Caryl Hill e Jan Hendry della Australian National University e C. David Ross, Mary Johnson e Bunge della Washington University School of Medicine hanno ottenuto risultati analoghi in animali integri. Quando piccoli frammenti di gangli simpatici venivano prelevati da ratti di varie età e trasferiti in coltura per un periodo di tempo standard, l'espressione della funzione colinergica declinava progressivamente con l'aumentare dell'età dell'animale da cui venivano prelevati i frammenti, raggiungendo un minimo nei gangli di ratti adulti.

Uno degli aspetti più interessanti dell'influenza esercitata dalle cellule diverse dai neuroni sulla scelta del neurotrasmettitore è la differente efficacia dimostrata da cellule prelevate in vari distretti corporei. Se questa proprietà venisse opportunamente localizzata, potrebbe agire da determinante specifico della scelta del trasmettitore. Per indagare su questo problema abbiamo studiato l'effetto di vari tessuti di ratto neonato sulla scelta del trasmettitore. Tutte le cellule diverse dai neuroni erano in grado di indurre la funzione colinergica, ma esistevano nette differenze quantitative nella loro efficacia: le più efficaci erano quelle della muscolatura scheletrica, in uno stato intermedio erano le cellule cardiache, mentre le meno efficaci erano quelle epatiche. Un notevole aspetto di questa gerarchia era il suo rapporto con la quantità di innervazione colinergica ricevuta dal tessuto bersaglio: le fibre dei muscoli scheletrici ricevono solo un input colinergico, le cellule della muscolatura cardiaca un miscuglio di input colinergico e adrenergico, infine le cellule epatiche, quando sono innervate, solo un input adrenergico. Questo risultato ha fatto pensare che, durante lo sviluppo, le cellule bersaglio abbiano liberato fattori chimici in grado di influenzare il metabolismo del trasmettitore nei neuroni che le innervano.

Tuttavia, altre osservazioni non sembrano poter rientrare in quest'ipotesi. Parecchi tipi di cellule che condizionano efficacemente il mezzo di coltura non sono innervati normalmente, non ricevono cioè sinapsi. Tra queste cellule si annoverano i fibroblasti (cellule del tessuto connettivo), una linea di cellule tumorali della glia coltivate in laboratorio (le cellule della glia sono le cellule di sostegno dei neuroni nell'encefalo e nel midollo spinale) e le cellule diverse dai neuroni che si trovano nei gangli simpatici. La capacità di queste ultime di indurre un differenziamento colinergico ci è parsa veramente paradossale: nelle prime tre settimane dopo la nascita tali cellule, nel ganglio, circondano ogni neurone intimamente; eppure la maggior parte dei neuroni presenti nei gangli intatti diventa adrenergica. Come mai le cellule diverse dai neuroni, o il terreno colturale condizionato, possono far diventare praticamente tutti i neuroni simpatici colinergici in coltura, mentre nell'organismo solo pochi di essi diventano tali anche in presenza di abbondanti cellule diverse dai neuroni?

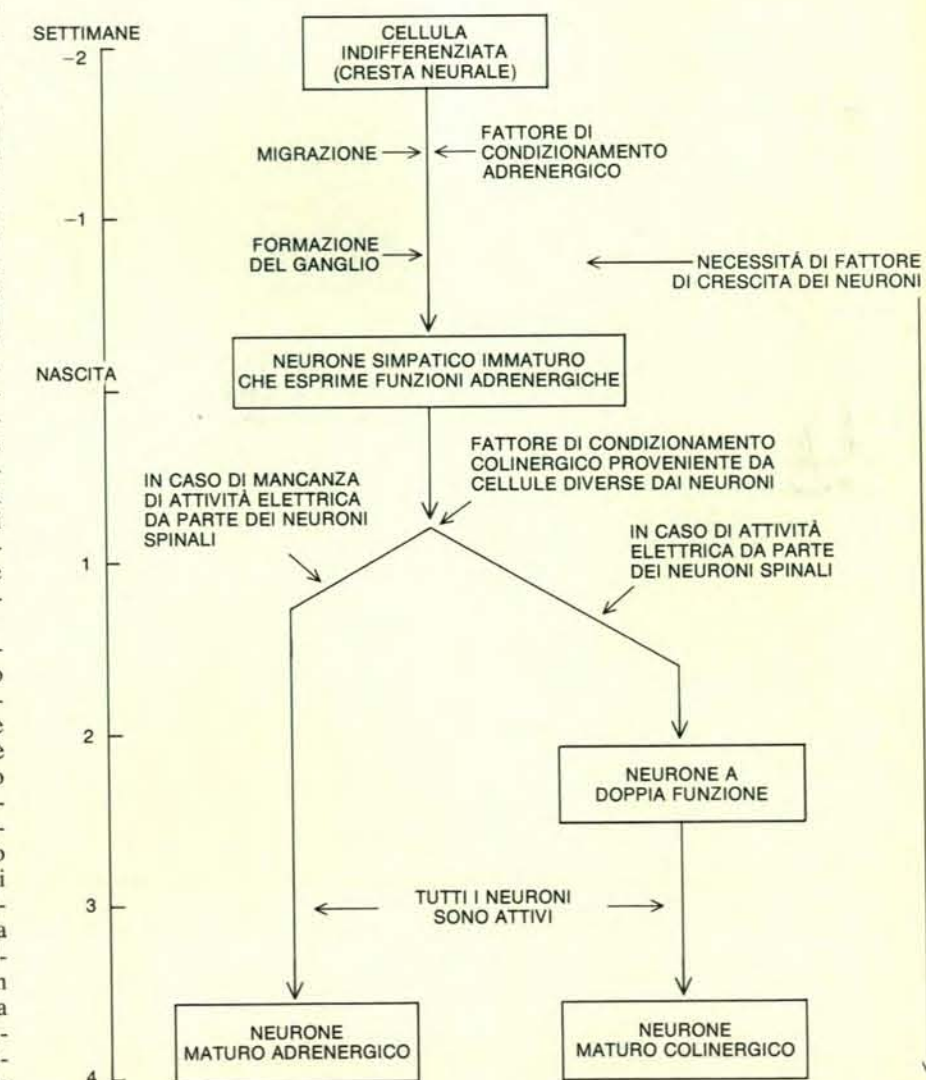
Questo evidente paradosso ha dimostrato che ulteriori variabili dovevano essere coinvolte nella scelta del trasmettitore. I neuroni simpatici nel corpo erano esposti al segnale colinergico lanciato dalle cellule diverse dai neuroni, ma la maggior parte di essi non poteva rispondervi per l'impedimento posto da qualche fattore che mancava, invece, nel terreno di coltura. Un possibile candidato a essere questo fattore era il NGF (*nerve growth factor*), dato che ha un'importanza critica per la sopravvivenza e la crescita dei neuroni simpatici adrenergici sia nell'animale integro sia in coltura. Inoltre, quando le colture di neuroni vengono esposte a elevate concentrazioni di NGF, si nota un aumento nella sintesi di componenti specificamente adrenergici, come la noradrenalina, rispetto alla sintesi di componenti non specifici della cellula, come le proteine o i lipidi. Risulta, però, che il NGF ha lo stesso effetto potenziatore sul differenziamento colinergico. Così, nei riguardi della produzione del trasmettitore, esso è più permissivo che istruttivo: stimola la crescita e il differenziamento dei neuroni simpatici immaturi lungo la via adrenergica o la via colinergica, mentre non ha alcuna influenza sulla scelta dell'una o dell'altra di queste vie. Contrariamente al NGF, il fattore colinergico secreto dalle cellule diverse dai neuroni non influisce sulla sopravvivenza o sulla crescita dei neuroni, mentre istruisce effettivamente i neuroni sulla scelta del neurotrasmettitore.

Un altro candidato a essere, nell'animale integro, il fattore a cui si è accennato prima, in grado di impedire alla grande parte dei neuroni simpatici di diventare colinergici, era il normale input eccitatorio da parte dei neuroni del sistema centrale, input assente nelle nostre colture. Ira Black, Hendry e Leslie Iversen del Laboratory of Molecular Biology del Medical Research Council di Cambridge hanno trovato che, interrompendo in giovani topi l'input spinale ai gangli simpatici, si riduceva l'ulteriore sviluppo del metabolismo adrenergico. Queste osservazioni hanno fatto formulare l'ipotesi che l'attività elettrica, imposta ai neuroni gangliari durante la prima settimana dopo la nascita, svolgesse un ruolo nel determinare la scelta del trasmettitore. Benché nella coltura non fossero presenti neuroni del sistema nervoso centrale, abbiamo potuto simulare il loro effetto eccitatorio sui neuroni gangliari (in cui la membrana risulta depolarizzata, cioè la tensione che l'attraversa risulta invertita di segno). Si è realizzato tutto questo aumentando la concentrazione del potassio nel mezzo ambientale, a cui si è aggiunto il farmaco veratridina, che provoca un afflusso di ioni sodio all'interno dei neuroni, o stimolando elettricamente le cellule una volta al secondo per molti giorni.

Quando colture di neuroni sono state depolarizzate o in presenza di un terreno condizionato o per un periodo di 7-10 giorni prima dell'aggiunta di tale mezzo, i neuroni sono rimasti prima di tutto adrenergici. In effetti, la depolarizzazione faceva abbassare il rapporto tra la sintesi dell'acetilcolina e la sintesi della noradrenalina di anche 300 volte, rispetto alle colture che ricevevano semplicemente il terreno condizionato. Questi mutamenti subentravano senza una significativa alterazione per quanto riguardava la sopravvivenza dei neuroni, suggerendo che questi ultimi, che avrebbero dovuto diventare colinergici in risposta al mezzo condizionato, rimanevano invece adrenergici. Era come se la depolarizzazione dei neuroni simpatici rendesse definitiva la loro istruzione prenatale a diventare adrenergici e limitasse, invece, la loro plasticità nei riguardi della scelta del trasmettitore.

L'attività elettrica si accompagna all'ingresso di ioni calcio nei neuroni. Dato che questo fatto è importante per controllare la secrezione dei neurotrasmettitori e probabilmente altre funzioni cellulari, è sorto il problema di sapere se si potevano diminuire gli effetti della depolarizzazione sulla scelta del trasmettitore, impedendo l'afflusso di calcio. Ciò è stato fatto aumentando la concentrazione degli ioni magnesio nel mezzo ambientale o aggiungendo la sostanza D600, che blocca selettivamente i movimenti degli ioni calcio attraverso la membrana cellulare. In effetti, quando l'ingresso del calcio veniva bloccato, la depolarizzazione non impediva più l'induzione delle proprietà colinergiche da parte del terreno colturale condizionato. Così, l'effetto della depolarizzazione nel bloccare l'influenza del mezzo condizionato sulla scelta del trasmettitore sembra essere mediato dagli ioni calcio, anche se non si riesce ancora a comprendere bene l'esatto meccanismo.

Queste osservazioni ci fanno ritenere che, nei gangli intatti, la maggioranza dei neuroni sia mantenuta nella condizione adrenergica prenatale da un input elettrico che proviene dal midollo spinale e che inizia a partire dalla prima settimana dopo la nascita. È anche possibile che i neuroni situati nei gangli destinati a diventare colinergici (quelli che innervano certi vasi sanguigni e certe ghiandole sudoripare) acquisiscano il loro input elettrico solo dopo essere stati influenzati da cellule diverse dai neuroni. Connessa con questo meccanismo vi è un'interessante possibilità: forse la formazione selettiva delle sinapsi tra i neuroni del sistema nervoso centrale e i neuroni dei gangli non solo stabilisce il tipo di circuito delle vie autonome, ma determina anche la scelta del trasmettitore adatto per quel tipo di circuito. Questo concetto sottolinea l'importanza potenziale dell'attività dei neuroni nel differenziamento chimico del sistema nervoso.

Gli esperimenti descritti in questo articolo lasciano scarsi dubbi sul fatto che i neuroni simpatici in via di sviluppo siano perlomeno provvisoriamente plastici nei riguardi della scelta del trasmettitore. Un'analoga conclusione ha tratto, di recente, Nicole Le Douarin assieme ai suoi collaboratori, presso l'Institut d'Embryologie di Nogent-sur-Marne (Francia), eseguendo ingegnosi esperimenti di trapianto in embrioni d'uccello. Essa ha trovato che la «decisione» dei neuroni immaturi di diventare adrenergici o colinergici può mutarsi se le cellule vengono trapiantate in nuove sedi nell'embrione. Per esempio, una popolazione di precursori, che normalmente dà origine a neuroni colinergici, produrrà neuroni adrenergici se viene trapiantata in una regione in cui normalmente compaiono neuroni di quest'ultimo tipo. In conclusione, sia in coltura sia negli embrioni integri, questa importante decisione da parte dei neuroni non è del tutto preprogrammata, ma può essere influenzata da altre cellule.

Il diagramma rappresenta un'ipotesi sulle fasi di sviluppo interessate nel differenziamento chimico di un neurone simpatico, secondo cui le influenze ambientali sulla scelta del neurotrasmettitore da parte della cellula vengono esercitate sia dalle cellule diverse dai neuroni (che liberano fattori di sviluppo) sia da altri neuroni (la cui attività elettrica modifica la risposta del neurone ai suddetti fattori). L'asse del tempo è approssimato; va ricordato che i neuroni maturano in tempi diversi.

Raggi cosmici di più alta energia

Particelle veloci provenienti dallo spazio, dotate di maggiore energia di quelle prodotte dagli acceleratori costruiti dall'uomo, sono rivelate dagli sciami di altre particelle che esse creano entrando nell'atmosfera

di John Linsley

In ogni istante cade sulla Terra una pioggia continua di particelle che si muovono quasi alla velocità della luce e provengono da tutte le direzioni. Queste particelle, i raggi cosmici, attraggono l'attenzione dei ricercatori da più di sei decenni. Se si studiassero i raggi cosmici sulla superficie della Luna, dove mancano sia un'atmosfera che l'influsso del campo magnetico terrestre, si troverebbe che essi sono costituiti principalmente dai nuclei degli elementi più abbondanti, quelli cioè compresi tra l'elemento di minor peso atomico, l'idrogeno, e il gruppo che comprende il ferro (peso atomico 56). Inoltre, si troverebbero qualche nucleo di elementi più pesanti, un po' di elettroni e di positoni, pochi raggi gamma e gli onnipresenti, ma elusivi, neutrini. Queste particelle sono chiamate raggi cosmici, perché si osservano in moto attraverso lo spazio esterno con una velocità prossima a quella della luce, cioè con una velocità pari a quella delle particelle che costituiscono il fascio di un sincrotone o di qualche altro acceleratore costruito dall'uomo. In effetti, alcune particelle dei raggi cosmici hanno un'energia molto più elevata di quella delle particelle veloci generate dai più grandi acceleratori costruiti finora. Sono questi raggi cosmici di altissima energia che ci lasciano più perplessi. Da dove vengono? In che modo sono accelerati fino a simili energie?

Sebbene i raggi cosmici di altissima energia siano particelle subatomiche, cioè i più piccoli oggetti conosciuti, le loro energie cinetiche sono così elevate da essere confrontabili con quelle degli oggetti con cui abbiamo a che fare nella vita di ogni giorno. I raggi cosmici più energetici osservati finora hanno energie pari a due o tre volte 10^{20} elettronvolt, pari cioè all'energia cinetica di una palla da tennis colpita molto bene.

La ricerca sui raggi cosmici nella regione di energia più elevata, cioè al di sopra di 10^{15} elettronvolt, presenta un aspetto particolare: infatti, cercando di risolvere problemi di astrofisica si incontrano continuamente problemi ancora insoluti che competono a una branca diversa della scienza, la fisica delle alte energie. Questa situazione era usuale in passato anche nella ricerca sui raggi cosmici di energia minore di 10^{15} elettronvolt. Passò molto tempo, dopo il volo in pallone con cui Victor F. Hess scoprì la radiazione cosmica, prima che si potesse rispondere alla domanda fondamentale: che cosa sono i raggi cosmici?

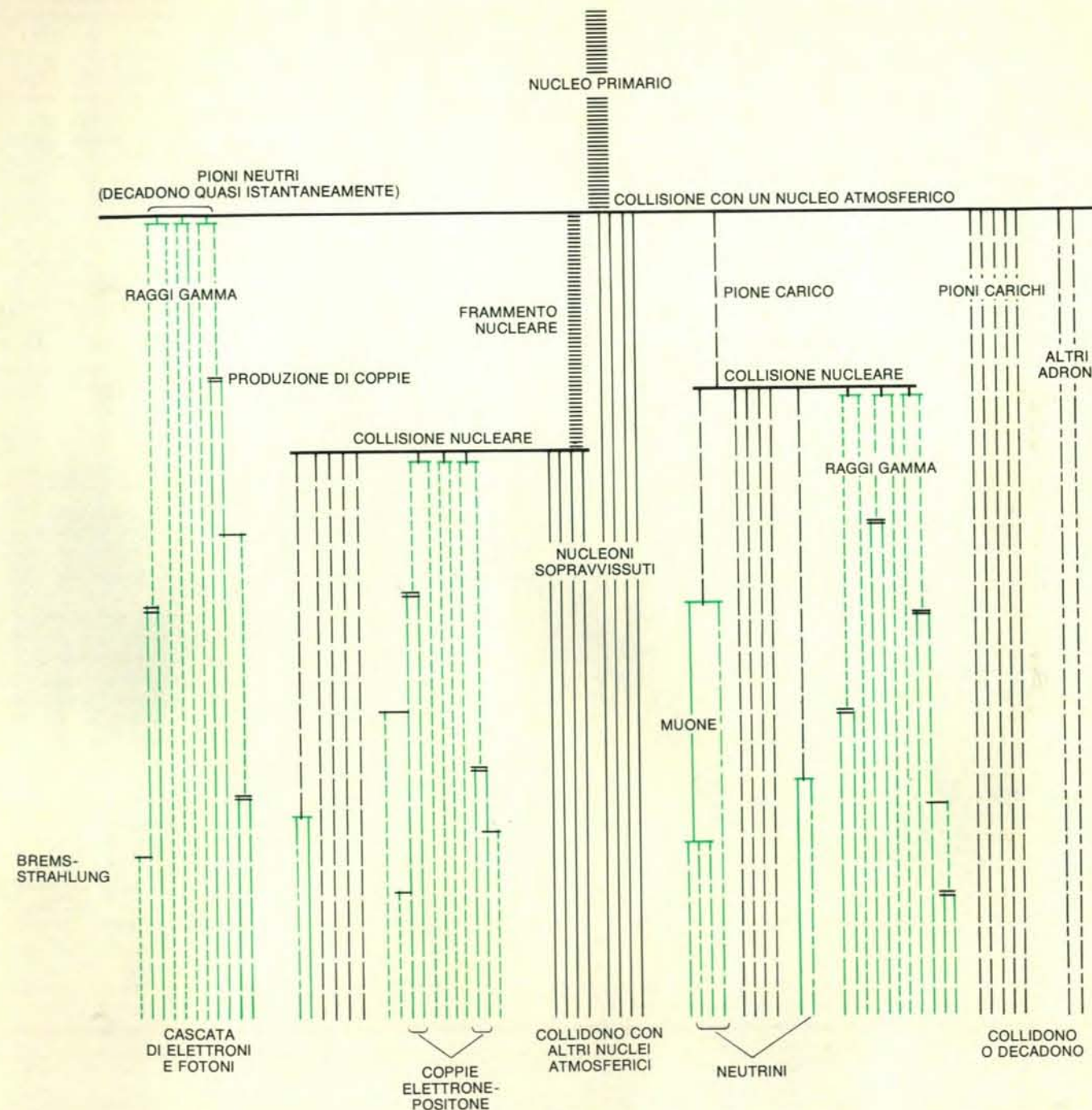
Dapprima i fisici cercarono di trovare la risposta usando le stesse tecniche che erano soliti utilizzare per identificare le particelle veloci emesse dai materiali radioattivi o quelle generate dai primi rudimentali «frantumatori di atomi». Incontrarono difficoltà. Trovarono, per esempio, che i raggi cosmici entrando nell'atmosfera praticamente mai rallentavano tranquillamente fino ad arrestarsi, come avrebbero fatto nel caso che le loro energie iniziali fossero state piccole come quelle delle particelle veloci che i ricercatori erano soliti identificare. Seguendo il percorso di ciò che chiamavano raggi cosmici primari trovarono invece che quasi tutte le traiettorie portavano alla sede di una violenta collisione tra la particella dei raggi cosmici e un nucleo atomico. Da lì avevano inizio le traiettorie di molte particelle subatomiche veloci dette raggi cosmici secondari.

I complessi fenomeni associati a questo tipo di collisioni si rivelarono così interessanti e importanti per una maggior comprensione della natura che lo studio di quei fenomeni fu portato avanti sempre più autonomamente e per sè solo. Quando si potè disporre di apparecchiature capaci di accelerare artificialmente elettroni e protoni fino a energie pari a quelle di almeno alcuni raggi cosmici, lo studio di questi fenomeni divenne indipendente e nacque la nuova branca chiamata fisica delle alte energie.

Non occorre dire che, se possibile, è molto meglio realizzare un dato esperimento di fisica delle alte energie utilizzando particelle accelerate artificialmente, piuttosto che usando raggi cosmici. Oggi si può riprodurre in laboratorio con gli acceleratori praticamente tutto ciò che accade in collisioni tra raggi cosmici e nuclei bersaglio fino a energie dei raggi cosmici di circa 10^{12} elettronvolt. Per il prossimo decennio sono programmate nuove macchine che porteranno questo valore fino a circa 10^{15} elettronvolt. I costi per la realizzazione di tali apparati sono però elevatissimi ed è possibile quindi che le nuove macchine siano le ultime della loro specie.

Il fatto che oggi si possano studiare collisioni come quelle tra raggi cosmici e nuclei atmosferici usando particelle accelerate artificialmente ha mutato la natura degli esperimenti con raggi cosmici nella zona di energia al di sotto di 10^{12} elettronvolt. Gli sperimentatori usano i dati accurati e dettagliati pubblicati dai fisici delle alte energie per calcolare la risposta dei propri strumenti alla radiazione cosmica. Oppure, semplicemente, trasportano i loro strumenti in laboratori dotati di acceleratori e li calibrano ponendoli in un fascio di particelle identiche alla componente di raggi cosmici che si sono prefissi di misurare. Per contro, un esperimento su raggi cosmici nella regione di energia al di sotto di 10^{12} elettronvolt avrà difficilmente lo scopo di acquisire nuovi dati sulle collisioni di alta energia.

Gli esperimenti nell'intervallo di energia compreso tra 10^{12} e 10^{15} elettronvolt hanno un carattere intermedio. Al di sopra di 10^{15} elettronvolt, nel dominio cioè dei raggi cosmici di altissima energia, essi hanno ancora il carattere che avevano in passato tutti gli esperimenti sui raggi cosmici: lavorando in questo settore capita infatti, per amore o per forza, di essere oggi astrofisici e domani fisici delle alte energie.

Indipendentemente dal punto di vista considerato, si incontrano sempre due difficoltà principali. Una è dovuta al fatto che l'atmosfera terrestre tende a impedire ai raggi cosmici primari di raggiungere il terreno. Per esempio, la probabilità che un protone cosmico raggiunga il livello del mare senza subire una collisione nell'atmosfera è solo un milionesimo circa. I raggi cosmici secondari prodotti nelle collisioni con nuclei atmosferici tendono a essere particelle instabili che decadono in altre particelle. A seguito di tutto questo la radiazione cosmica che è presente al livello del terreno è composta quasi esclusivamente dei prodotti di decadimento: muoni, neutrini e raggi gamma, insieme con elettroni, che sono il prodotto secondario dei muoni e dei raggi gamma. Se, come astrofisici, si vuole studiare la radiazione primaria direttamente, si devono mandare rivelatori al di sopra dell'atmosfera con palloni stratosferici, razzi e sa-

Uno sciame atmosferico si genera quando un nucleo dei raggi cosmici primari, per esempio una particella alfa (un nucleo di elio), entra nell'atmosfera terrestre circa 20 chilometri sopra la superficie e urta il nucleo di un atomo dell'atmosfera. Questo diagramma molto schematizzato mostra le prime fasi dello sviluppo dello sciame. Le linee in colore rappresentano leptoni: cioè particelle come fotoni (che comprendono i raggi gamma), neutrini, elettroni, positoni e muoni. Le linee in nero rappresentano adroni: particelle più pesanti, come nucleoni (protoni e neutroni), nuclei compositi e pioni. Le particelle prodotte nelle due collisioni secondarie rappresentate nella figura danno luogo alla seconda generazione di particelle. Le particelle eventualmente prodotte da queste ultime, nel caso che urtino altre particelle prima di decadere, costituiscono la terza generazione, e così via. Gli adroni possono generare sia leptoni che altri adroni, a differenza dei leptoni, che possono produrre solo altri leptoni. La dicitura «produzione di coppie» sta a indicare la materializzazione di coppie elettrone-positrone da un fotone. Il termine «Bremsstrahlung» (radiazione di frenamento) si riferisce a fotoni generati dalla decelerazione di altre particelle.

telliti artificiali. Se, come fisici delle alte energie, si vogliono studiare le collisioni tra nuclei di alta energia e nuclei atomici di un bersaglio, può essere necessario fare la stessa cosa, o almeno andare a lavorare in alta montagna.

Una difficoltà ancora maggiore è posta dalla bassa intensità della radiazione cosmica. Anche nell'intervallo di energia più favorevole (tra 10^9 e 10^{10} elettronvolt) e nel luogo più favorevole (al di sopra dell'atmosfera terrestre) l'intensità è appena sufficiente a produrre 10 urti al minuto su un'area di un centimetro quadrato. Questo tasso è così piccolo che nonostante la grande energia di cui sono dotate singolarmente le particelle dei raggi cosmici, l'energia totale che essi trasmettono nell'unità di tempo a un simile bersaglio è analoga a quella ricevuta sotto forma di luce stellare. Inoltre, l'intensità di quelli che chiamo qui «raggi cosmici ad altissima energia» è di gran lunga minore. La frazione di raggi cosmici primari di energia superiore a 10^{15} elettronvolt è così piccola che occorrerebbe un rivelatore di almeno 10 metri di lato per catturare una particella in un'ora. Naturalmente, indipendentemente dalle dimensioni non esiste strumento capace di catturare (in senso letterale) una simile particella con tutta la sua energia intatta.

Come fecero Pierre Auger e i suoi colleghi dell'École Normale Supérieure di Parigi a prevedere nel 1938 l'esistenza di raggi cosmici primari di energia fino a 10^{15} elettronvolt, facendo uso dei dati ottenuti con pochi contatori Geiger che avevano complessivamente un'area inferiore a un metro quadrato? Giunsero a tale conclusione mostrando che le scariche simultanee dei contatori osservate erano il risultato di un fenomeno precedentemente sconosciuto connesso ai raggi cosmici, e dimostrando alcuni aspetti essenziali del fenomeno. Attribuirono le scariche simultanee a sciami di particelle di raggi cosmici che cadevano sui loro strumenti provenendo dall'aria sopra di essi. Dal fatto che si osservavano conteggi simultanei anche quando i loro contatori Geiger erano separati da molte decine di metri dedussero che gli sciami erano estesi e coprivano un'area di qualche centinaia di metri quadrati, e che il numero di particelle veloci presenti negli sciami era grande, pari forse a decine di migliaia.

Gli sciami scoperti da Auger e dai suoi colleghi sono detti sciami estesi atmosferici o semplicemente sciami atmosferici. Non si deve però prendere il nome troppo alla lettera. Gli sciami di raggi cosmici hanno una struttura complessa, anche se ben definita, e iniziano e finiscono in un batter d'occhio. Rivelando gli sciami atmosferici invece delle particelle primarie che li producono si possono superare entrambe le difficoltà già descritte. Gli strumenti destinati allo studio degli sciami possono essere posti al livello del terreno; non è necessario portarli al di sopra dell'atmosfera. Inoltre, l'apparato non deve coprire completamente una grande area. Dato che gli sciami sono estesi, possono essere rivelati a una certa distanza. Si può far sì che una vasta area sia sensibile all'impatto di sciami atmosferici disponendo su di essa una schiera di piccoli rivelatori per particelle veloci ampiamente spaziati; in tal modo si otterranno dati con un tasso ragionevole nonostante che la intensità da misurarsi sia molto bassa.

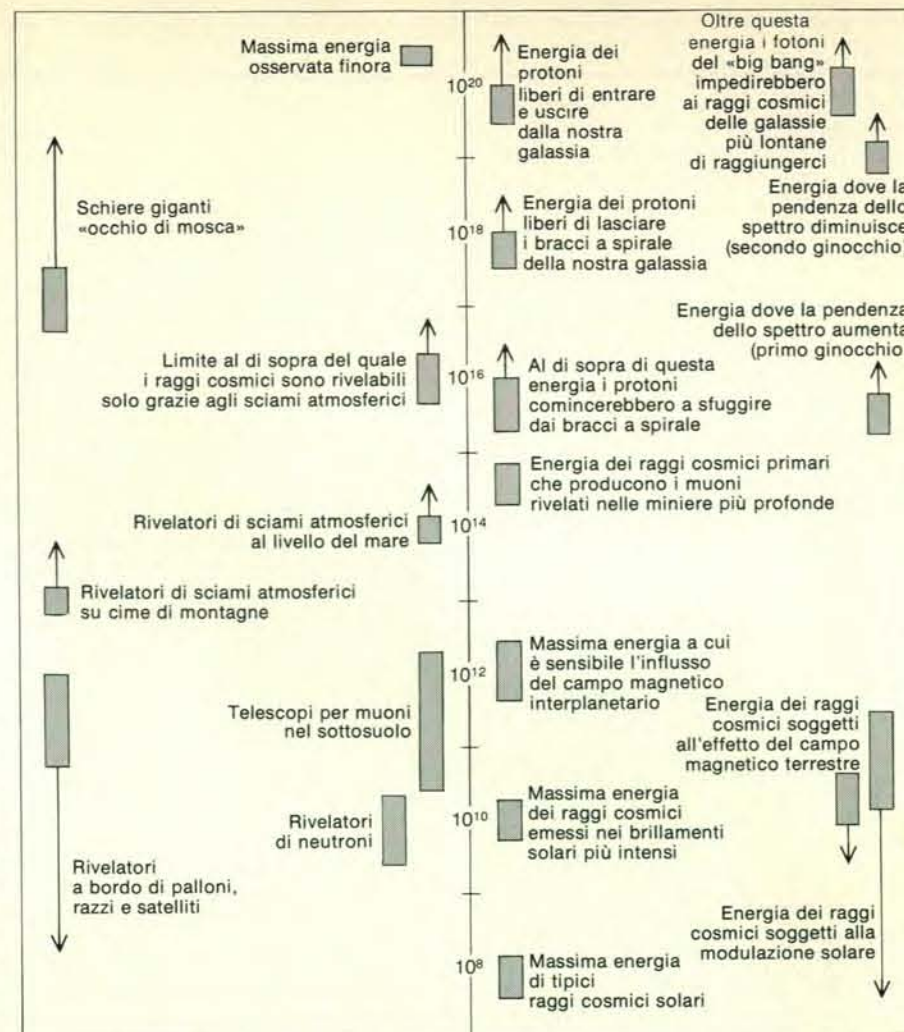
L'anno prima di quello in cui Auger e i suoi collaboratori scoprirono gli sciami atmosferici era stata sviluppata quasi contemporaneamente da Homi J. Bhabha e Walter F. Heitler in Inghilterra e da J. F. Carlson e J. Robert Oppenheimer negli USA una spiegazione teorica di sciami analoghi prodotti da raggi cosmici in bersagli solidi. Cercando di conservare quell'interpretazione, si pensò dapprima che

Sciami atmosferici sono rivelati da una schiera gigante di rivelatori indicati dai puntini su questa fotografia aerea di Haverah Park nei pressi di Leeds in Inghilterra. Il diametro della zona grossolanamente esagonale coperta dai rivelatori è di quattro chilometri circa. Ogni rivelatore è costituito da vari recipienti sigillati contenenti acqua purissima con sensori (tubi fotomoltiplicatori), che sono attivati dalla luce Čerenkov generata dai raggi cosmici secondari che penetrano nell'acqua. L'area totale dei 32 rivelatori indicati è di 500 metri quadrati.

gli sciami atmosferici fossero semplicemente cascate di elettroni e fotoni iniziate nell'alta atmosfera da elettroni primari dei raggi cosmici. Una gran mole di dati sperimentali mostra ora che quasi tutti i raggi cosmici che generano sciami atmosferici sono nuclei atomici, e che gli sciami atmosferici sono molto più complessi di quanto si era immaginato in un primo tempo.

Nella parte centrale di uno sciame atmosferico si trova una cascata nucleare connessa a una catena di interazioni nucleari. Lo sciame ha inizio con una collisione tra due nuclei atomici, la particella cosmica primaria e il nucleo di un atomo atmosferico. Molti raggi cosmici secondari prodotti in quella collisione, ma non tutti, subiscono collisioni analoghe più avanti lungo l'asse della cascata, producendo a loro volta particelle secondarie, e così via. In ogni generazione successiva prodotta dalle collisioni si ha però il decadimento di alcuni raggi cosmici secondari, soprattutto pioni.

I pioni carichi prodotti nelle prime fasi della cascata urtano solitamente nuclei di atomi atmosferici invece di decadere, dato che le loro vite medie sono abbastanza lunghe, se si tiene conto della dilatazione temporale prevista dalla teoria della relatività ristretta. Quelli prodotti nelle ultime fasi delle cascate solitamente decadono, dato che le loro energie sono minori e pertanto l'effetto relativistico non è così grande. Le particelle prodotte nel decadimento dei pioni carichi sono muoni e neutrini. I pioni neutri sono così instabili che decadono subito, generando coppie di fotoni di alta energia. Questi fotoni, che ricevono una frazione significativa dell'energia liberata in ogni collisione, danno inizio alle cascate di elettroni e fotoni. Riassumendo, uno sciame atmosferico di raggi cosmici è costituito da una cascata di nucleoni (protoni e neutroni) nella sua parte centrale, e da due gruppi di particelle subatomiche veloci provenienti da processi di decadimento: una componente di muoni e neutrini e un'altra di elettroni e fotoni.

La componente di elettroni e fotoni (elettromagnetica) è costituita da un gran numero di cascate di elettroni e fotoni. Tali cascate implicano anche una catena di collisioni. Sono collisioni però in cui un elettrone o un fotone di alta energia interagiscono col campo elettrico di un atomo. L'interazione forte e a breve raggio d'azione che tiene insieme i nucleoni formando i nuclei atomici non gioca nessun ruolo. Nelle prime fasi di una cascata di elettroni e fotoni il numero di particelle veloci cresce, sia perché un fotone di alta energia incontrando un campo elettrico intenso può materializzarsi, formando due particelle, un elettrone e un positone, sia perché un elettrone o un positone di alta energia, incontrando un campo elettrico intenso, possono irradiare una frazione significativa della propria energia sotto forma di un fotone molto energetico. Al crescere del numero delle particelle veloci deve diminuire, però, l'energia media per particella. Alla fine entrano in gioco processi dissipativi e la cascata si esaurisce.

L'energia dei raggi cosmici osservati varia entro l'enorme intervallo compreso tra 10^9 e 10^{20} elettronvolt *(a destra)*. Bisogna tener conto degli intervalli di energia e delle limitazioni presentate dai vari tipi di rivelatori per studiare specie diverse di raggi cosmici *(a sinistra)*. La lunghezza dei rettangoli indica, sia per le particelle che per i rivelatori, l'intervallo di energia tipico. Le frecce rivolte verso l'alto stanno a indicare che l'intervallo rappresentato costituisce un limite inferiore, le frecce dirette verso il basso specificano che l'intervallo costituisce invece un limite superiore.

Uno sciame atmosferico può essere visualizzato costruendo un modellino in scala 1:20 000 costituito da un disco di mezzo centimetro di diametro ritagliato da carta spessa e fatto scorrere su un filo lungo un metro. Il filo rappresenta l'asse dello sciame, prolungamento della traiettoria della particella primaria. Se si tiene il filo verticale, l'estremità superiore corrisponderà a un'altezza tipica per la prima collisione (nell'ipotesi che l'estremità inferiore sia al livello del mare). Per farsi un'idea della brevità della vita di uno sciame atmosferico si immagini che il disco scorra giù lungo il filo alla velocità della luce.

Il disco rappresenta la regione in cui ci si aspetta di trovare gli elettroni. Essi costituiscono la componente di particelle cariche di gran lunga più abbondante negli sciami atmosferici. La regione centrale, in cui ha sede la cascata nucleonica, è grande solo quanto il foro attraverso il quale passa il filo: negli sciami atmosferici reali ha un diametro di pochi metri. Il disco rappresenta con buona approssimazione la distribuzione longitudinale degli elettroni. Nella zona centrale degli sciami atmosferici reali la dispersione longitudinale degli elettroni è solo di un metro circa. Perpendicolarmente all'asse il disco ci dà una rappresentazione più semplice della situazione reale. A differenza della densità della carta, che è costante fino a un certo raggio per poi cadere a zero bruscamente, la densità degli elettroni negli sciami decresce regolarmente e senza discontinuità al crescere della distanza dall'asse, senza che si siano mai osservate brusche interruzioni. Il disco del modello rappresenta la regione che contiene metà degli elettroni. Negli sciami atmosferici reali il diametro di questa regione è pari a 100 metri circa, indipendentemente dall'energia della particella che ha dato inizio allo sciame. Comunque, la densità del-

le particelle nello sciame cresce al crescere dell'energia primaria a qualsiasi distanza dall'asse dello sciame, così che la densità delle particelle in sciami altamente energetici è così elevata da rendere rivelabile lo sciame stesso anche a un chilometro di distanza dall'asse con contatori di dimensioni modeste.

Se si potesse seguire uno sciame atmosferico nel suo moto e osservarne lo sviluppo, si vedrebbe che il numero di particelle in un primo tempo cresce rapidamente, raggiungendo un massimo in cui la creazione di nuove particelle è controbilanciata dall'assorbimento, e poi diminuisce. Il numero delle particelle che compongono lo sciame nella fase detta di «massimo sviluppo», cioè il massimo numero di particelle presente in un dato sciame atmosferico, è proporzionale all'energia della particella cosmica primaria che ha generato lo sciame. Per esempio, un raggio cosmico primario dotato di un'energia di 10^{15} elettronvolt genererà uno sciame atmosferico contenente mezzo milione di particelle al massimo sviluppo. Questi sciami relativamente poco energetici raggiungono il massimo sviluppo nell'alta atmosfera. Uno sciame di tale energia, però, può ancora essere rivelato facilmente al livello del mare nel caso che il raggio cosmico primario stesse viaggiando in una direzione quasi verticale al suo ingresso nell'atmosfera. Gli sciami atmosferici prodotti dai raggi cosmici più energetici osservati finora, dotati cioè di energia pari a due o tre volte 10^{20} elettronvolt, sembrano raggiungere il massimo sviluppo dopo avere attraversato circa otto decimi della atmosfera, cioè a una quota di circa 2000 metri sopra il livello del mare.

Gli esperimenti di Auger e dei suoi collaboratori fornirono un metodo per studiare i raggi cosmici di altissima energia, mostrando che gli sciami atmosferici permettono di compensare la rapida diminuzione dell'intensità osservata dei raggi cosmici al crescere dell'energia. In quegli esperimenti, però, ci si limitò a rivelare gli sciami, senza effettuare su di essi misurazioni particolari. È evidente che non è possibile determinare il numero e la direzione di provenienza dei raggi cosmici che generano gli sciami avendo come unica informazione il fatto che sono state osservate scariche simultanee in contatori Geiger ampiamente distanziati. La conclusione raggiunta da Auger e dai suoi collaboratori, cioè l'esistenza di sciami atmosferici, alcuni dei quali di energia superiore a 10^{15} elettronvolt, si basava sull'analisi statistica di molti eventi, con l'ausilio della teoria delle cascate di elettroni e fotoni da poco sviluppata.

Nel 1948 si ebbe un importante progresso nello studio degli sciami atmosferici per opera di Robert W. Williams del Massachusetts Institute of Technology. Invece di usare tubi Geiger egli utilizzò camere a ionizzazione, di cui registrava i segnali fotografando tracce di oscilloscopio. Lavorando sul monte Evans nel Colorado, egli collocò i suoi rivelatori al centro e ai vertici di un triangolo equilatero di 12 metri di lato. Quando uno sciame atmosferico abbastanza grande cadeva all'interno della sua schiera di rivelatori o poco fuori di essa, le particelle veloci producevano segnali praticamente simultanei in tutte le camere a ionizzazione, così che l'evento poteva essere riconosciuto nel solito modo. Le tracce fotografate da Williams sugli oscilloscopi gli fornivano, però, informazioni più dettagliate del semplice dato che uno sciame aveva colpito il suo apparato. Egli sapeva che l'ampiezza del segnale proveniente da ogni camera a ionizzazione era proporzionale al numero di particelle che erano passate attraverso di essa. Disponendo di questi dati, Williams potè localizzare l'asse dello sciame ed effettuare un calcolo del numero di particelle presenti nello sciame stesso.

Cinque anni dopo, nel 1953, la metodologia per le osservazioni degli sciami atmosferici fu perfezionata ulteriormente da Pietro Bassi del MIT, che lavorava con George W. Clark e Bruno Rossi. Invece di camere a ionizzazione usarono contatori a scintillazione, distanziati tra loro di 30 metri e connessi con un sistema elettronico di registrazione capace di una risposta molto rapida. Furono loro che mostrarono, lavorando con questo apparato, che le particelle in prossimità dell'asse di uno sciame atmosferico si muovono come un disco piatto e sottile. Riuscirono inoltre a calcolare la direzione di movimento di singoli sciami deducendola dal ritardo temporale che osservavano tra coppie di rivelatori posti su linee di riferimento perpendicolari tra loro.

I contatori a scintillazione usati da Bassi e dai suoi collaboratori sono simili alle camere a ionizzazione nel fatto che l'ampiezza del segnale in uscita è proporzionale al numero di particelle veloci che generano il segnale. Questo aspetto non fu utilizzato nell'esperimento che ho appena descritto, ma non si trascurò la possibilità di farne uso in un esperimento successivo. Quasi subito si formò al MIT un gruppo più ampio sotto la guida di Rossi. Questo gruppo si proponeva di combinare la nuova tecnica per determinare la direzione di provenienza col metodo usato in precedenza da Williams per valutare il numero di particelle negli sciami. Già nel 1956 il gruppo di Rossi stava effettuando misure di sciami atmosferici da raggi cosmici con una schiera di 15 contatori a scintillazione molto più grandi, nei pressi di Boston, all'Agassiz Station dello Harvard College Observatory. L'area della schiera era pari a circa un sesto di chilometro quadrato. Si poteva determinare il numero di particelle con un'accuratezza del 20 per cento circa e si poteva misurare la direzione degli sciami con un'incertezza di pochi gradi. La tecnica di Williams-Bassi fu presto adottata da quasi tutti gli altri gruppi di ricerca che si occupavano di sciami atmosferici.

Mappa di intensità di uno sciame atmosferico di grandi dimensioni prodotto da un raggio cosmico con un'energia di 10^{19} elettronvolt registrato nel marzo 1973 da una schiera di rivelatori a Volcano Ranch nel New Mexico. Le linee concentriche in colore rappresentano contorni di uguale intensità misurata in termini di numero di particelle per metro quadrato. L'asse dello sciame era quasi verticale, così che le curve di uguale intensità sono quasi circolari. I cerchi neri rappresentano contatori a scintillazione del diametro di un metro posti sul terreno e distanziati di 150 metri. (I rivelatori non sono stati riportati in scala: altrimenti sarebbero stati più piccoli del «punto» alla fine di questa frase.) Il numero all'interno di ogni rivelatore rappresenta il numero di particelle dello sciame misurato in quel punto. I segnali di maggiore ampiezza indicano che l'asse dello sciame era a pochi metri dal centro di un triangolo formato dai rivelatori in colore più intenso.

Il risultato più significativo ottenuto con la schiera di Agassiz fu l'acquisizione di prove sperimentali, fornite nel 1957 da uno sciame atmosferico insolitamente esteso, che l'energia dei raggi cosmici primari poteva raggiungere i 5×10^{18} elettronvolt. Questa scoperta creò agitazione nella comunità astrofisica a causa delle considerazioni che si potevano fare relativamente al problema dell'origine dei raggi cosmici. Proprio pochi anni prima i primi dati sugli sciami atmosferici avevano fatto cadere un'ipotesi avanzata nel 1949 da R. D. Richtmeyer e da Edward Teller: che tutti i raggi cosmici potessero provenire dal Sole. Richtmeyer e Teller ritenevano che l'isotropia osservata nei raggi cosmici (cioè il fatto che presentino, fuori dall'influsso terrestre, la stessa intensità in tutte le direzioni) potesse essere spiegata come il risultato dell'azione di un ipotetico campo magnetico che li confina (piegandone le traiettorie) entro un grande volume contenente il Sole. Essi ammettevano che non sarebbe stato possibile confinare particelle di energia molto maggiore di 10^{16} elettronvolt, a meno di ipotizzare un volume troppo grande o un campo magnetico troppo intenso. Quasi immediatamente altri ricercatori dedussero dagli sciami atmosferici l'esistenza di molti raggi cosmici primari di energia maggiore di 10^{16} elettronvolt, escludendo così l'ipotesi che il Sole fosse la sorgente di tutti i raggi cosmici.

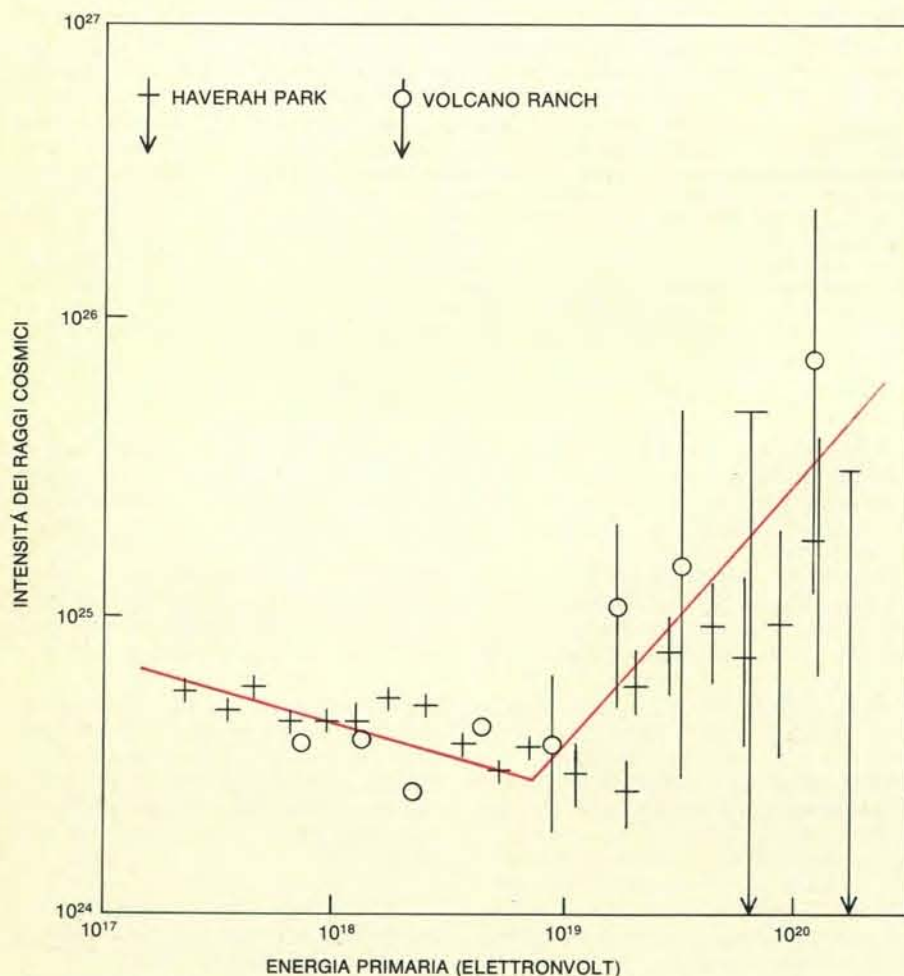
Anche se l'ipotesi dell'origine solare non ha più credito, argomentazioni simili a quelle usate da Richtmeyer e Teller continuano a essere citate quando si discutono alternative all'ipotesi dell'origine solare. Per esempio, l'idea che i raggi cosmici siano confinati per effetto di un campo magnetico può essere applicata all'intera Galassia. Anche nell'ipotesi dell'origine galattica c'è un'energia critica per i raggi cosmici, analoga a quella di 10^{16} elettronvolt prevista per l'ipotesi solare di Richtmeyer e Teller. Il campo magnetico della Galassia sembra essere un po' più debole del campo ipotizzato da Richtmeyer e Teller, ma il volume della Galassia è molto maggiore del volume proposto da loro. Pertanto l'energia critica dei raggi cosmici nell'ipotesi del confinamento galattico supera 10^{16} elettronvolt di numerosi ordini di grandezza.

Il nuovo record di energia per un raggio cosmico stabilito dallo sciame atmosferico più grande osservato con la schiera di Agassiz non superava il «limite galattico», ma risultava scomodamente vicino a tale limite per i sostenitori della teoria dell'origine galattica. La schiera dell'Agassiz Station era entrata da poco in funzione che già alla Cornell University aveva inizio il lavoro per realizzare una schiera analoga che avrebbe avuto un'area sensibile quasi cinque volte maggiore. Al MIT si facevano piani per ridislocare in Bolivia i rivelatori dell'Agassiz Station così da formare una schiera con un'area pari a quattro volte quella originale. I risultati raccolti nell'esperimento di Agassiz Station confermavano, però, che l'intensità della radiazione cosmica diminuisce molto rapidamente al crescere dell'energia. Pertanto, schiere che fossero state soltanto quattro o cinque volte più grandi di quella realizzata ad Agassiz Station non avrebbero potuto rispondere alla domanda se i raggi cosmici siano o meno tanto energetici da non poter essere contenuti nella Galassia.

Come nuovo membro del gruppo di Rossi vagheggiavo privatamente programmi grandiosi sulla schiera più grande che ritenevo realistico costruire e far funzionare. Sarebbe stata gigantesca, non cinque, ma 50 volte più grande della schiera di Agassiz Station. Non ero l'unico del nostro gruppo che concepiva simili sogni, ma ero l'unico a quel tempo abbastanza libero per cercare di portare a compimento il progetto. Avevo l'incoraggiamento di Rossi, che fece in modo che io potessi disporre di tutte le risorse materiali necessarie. Soprattutto, fu attraverso Rossi che ottenni un inestimabile collaboratore, Livio Scarsi, dell'Università di Milano, ma temporaneamente al MIT. A partire dal 1959 Scarsi e io realizzammo la nostra schiera gigante a Volcano Ranch presso Albuquerque nel New Mexico.

Il ghiaccio era stato rotto: si era visto che le schiere giganti costituiscono una possibilità realistica. In pochi anni un gruppo guidato da J. G. Wilson dell'Università di Leeds realizzò e fece funzionare una seconda schiera gigante a Haverah Park presso Leeds. In seguito furono allestite altre due schiere giganti, una in Australia da un gruppo dell'Università di Sidney guidato da C.B.A. McCusker e l'altra in Siberia presso la città di Ya-

Mappa dei tempi di arrivo dello sciame atmosferico del marzo 1973, ottenuta anch'essa con rivelatori della schiera di Volcano Ranch. (Per ridurre i costi non tutti i rivelatori erano equipaggiati in modo da dare misure di tempi di arrivo oltre che di intensità.) Le linee diagonali in colore indicano il luogo ove il piano del disco dello sciame intersecò il piano del terreno in istanti di tempo successivi. Dato che l'asse dello sciame era quasi verticale, il piano dello sciame era quasi orizzontale, e la linea della sua intersezione col terreno passò attraverso la schiera di rivelatori in meno di un microsecondo (un milionesimo di secondo). L'orientazione dell'asse fu determinata dai tempi di arrivo misurati (*numeri in nero*), adattando a essi un piano ipotetico. L'accordo tra i tempi di arrivo teorico e quelli misurati ci dice che la direzione dello sciame è stata determinata con un'incertezza di pochi gradi. I tempi sono in microsecondi da un istante iniziale arbitrario.

Le schiere giganti sono indicate di solito col nome della località ove sono poste. Non esiste una sistemazione ottimale per disporre i rivelatori in una schiera. Da sinistra a destra sono qui schematizzate le schiere di Volcano Ranch, di Sidney in Australia, di Yakutsk in URSS. Il cerchio in colore, che rappresenta uno sciame atmosferico con un diametro di due chilometri, dà un'idea delle dimensioni delle schiere. La schiera di Volcano Ranch non ha più l'assetto rappresentato qui sopra. Dopo quattro anni di funzionamento (dal 1959 al 1963) fu ridisposta formando la schiera più piccola, ma più fitta, rappresentata nelle illustrazioni delle due pagine precedenti.

Lo spettro di energia dei raggi cosmici primari di energia più elevata è stato determinato con i dati forniti da due schiere giganti. Sull'asse verticale è riportata l'intensità dei raggi cosmici in termini della funzione di Hillas (l'intensità differenziale moltiplicata per il cubo dell'energia). Se lo spettro avesse semplicemente la forma di una legge di potenza cubica inversa, i punti sperimentali si disporrebbero attorno a una linea orizzontale. In realtà, lo spettro appare un po' più ripido di una legge cubica inversa fino a un'energia di 10^{19} elettronvolt, per diventare poi più piatto a energie superiori. La zona dello spettro di energia dove lo spettro cambia forma è detta «ginocchio».

kutsk, pochi gradi a sud del circolo polare artico, per opera di un gruppo appartenente all'Istituto di ricerca di fisica cosmica e aeronomia diretto da Yu. G. Shafer.

L'ipotesi che i raggi cosmici siano confinati all'interno delle galassie era stata messa in forse nel 1957 dagli sciami atmosferici più energetici osservati con la schiera dell'Agassiz Station. Nel 1962 tale ipotesi subì un duro colpo ad opera del più grande sciame atmosferico che Scarsi e io osservammo a Volcano Ranch, uno sciame per cui si stimò una energia 10^{20} elettronvolt. Quando Scarsi e io intraprendemmo il nostro progetto non ci aspettavamo di rivelare uno sciame così energetico. Altrettanto inattesa fu la nostra scoperta che il numero di particelle primarie di energia superiore a qualche 10^{18}elettronvolt era più elevato del previsto. In corrispondenza a quelle energie tanto elevate lo spettro dei raggi cosmici diventava più piatto: presentava un «ginocchio». (Lo spettro di energia dei raggi cosmici è la curva che dà la variazione dell'intensità dei raggi cosmici man mano che ci si sposta in su o in giù lungo la scala delle energie.)

L'appiattimento dello spettro indicato dai dati di Volcano Ranch era un risultato inatteso per tutti. In effetti, i sostenitori della teoria dell'origine galattica avevano previsto proprio il contrario. Essi prevedevano che lo spettro di energia dei raggi cosmici sarebbe diventato più ripido alle energie che si approssimano all'energia critica per il confinamento galattico. È possibile che al di sopra di una certa energia la perdita di raggi cosmici dalla nostra Galassia sia più che compensata da raggi cosmici che entrano nella Galassia provenendo dallo spazio intergalattico? Se è così, qual è la sorgente dei raggi cosmici che entrano? I quasar o altri oggetti noti all'astronomia sarebbero capaci di riempire gli enormi spazi tra le galassie con particelle subatomiche dotate di energia tanto elevata?

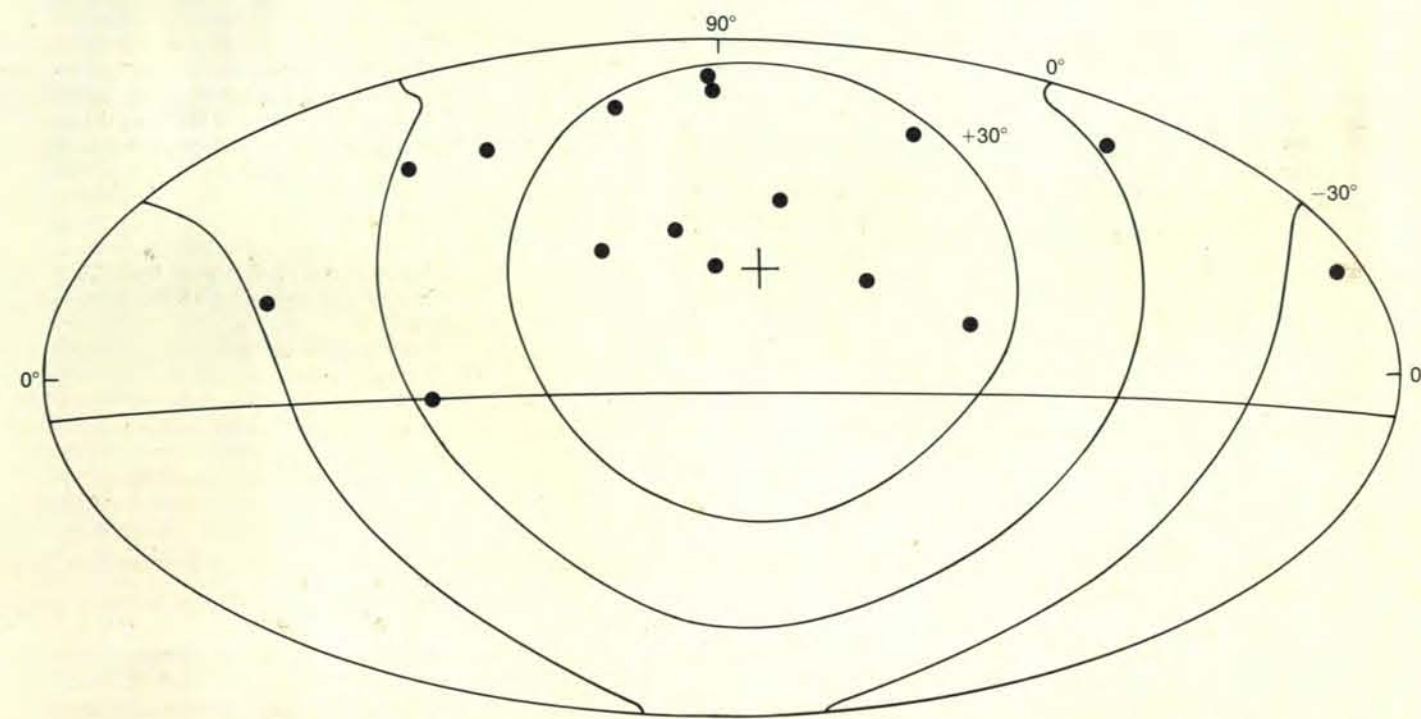
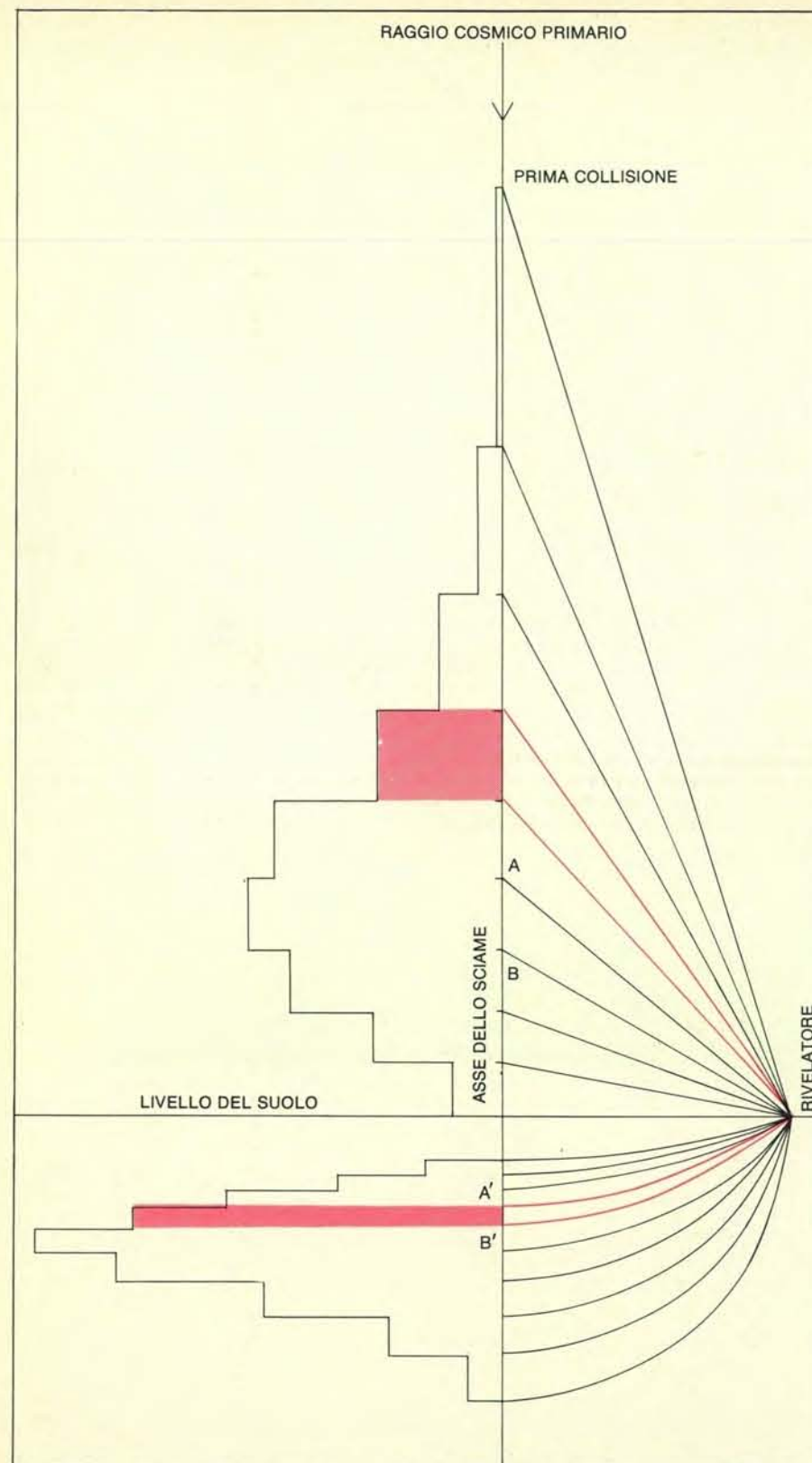
L'evento da 10^{20} elettronvolt osservato a Volcano Ranch assunse presto ulteriore importanza. Nel 1965 Arno A. Penzias e Robert W. Wilson dei Bell Laboratories scoprirono nelle microonde una particolare radiazione elettromagnetica che sembra permeare l'intero universo. Essa è detta talvolta «radiazione residua» dato che si pensa che sia la radiazione lasciata a seguito del «big bang», con cui ebbe inizio l'universo. Venendo a conoscenza della scoperta di Penzias e Wilson, Kenneth Greisen di Cornell e G. T. Zatsepin dell'Istituto di fisica P. N. Lebedev di Mosca pensarono indipendentemente che le interazioni tra i raggi cosmici e la radiazione residua potessero fare da filtro per i raggi cosmici di energia superiore a una certa energia di soglia facilmente calcolabile. Dimostrarono che, se le sorgenti di raggi cosmici sono distribuite uniformemente nell'universo, allora, a seguito di tali interazioni, lo spettro dei raggi cosmici dovrebbe diventare più ripido bruscamente a circa 6×10^{19} elettronvolt. L'intensità dei raggi cosmici di energia pari a 10^{20} elettronvolt sarebbe così bassa da rendere impossibile l'osservazione di sciami di 10^{20} elettronvolt.

Avevamo forse sbagliato Scarsi e io nel valutare l'energia del nostro sciame più esteso? Sembrava di no. Quando le altre schiere giganti incominciarono a funzionare registrarono anch'esse sciami di energia pari a quella del nostro evento e anche maggiori. Anche la diminuzione di pendenza nello spettro è stata confermata. Quale quadro suggeriscono questi risultati per l'origine dei raggi cosmici? E quali altri dati si possono ricavare dalle schiere giganti?

Si ritiene in genere che praticamente tutti i raggi cosmici di energia inferiore al valore per cui lo spettro cambia pendenza siano generati da sorgenti interne alla nostra Galassia. Inoltre, è conveniente supporre che questi raggi cosmici siano confinati nella Galassia, perché in tal caso non è necessario che le ipotetiche sorgenti siano particolarmente potent i. Se i raggi cosmici sono effettivamente confinati, la loro intensità nello spazio interstellare cresce fino a a stabilire un bilancio con la dispersione nello spazio intergalattico (le perdite all'interno della Galassia non sono rilevanti). Dato che il confinamento ipotizzato è dovuto ai campi magnetici, esso non è ugualmente efficace su raggi cosmici di energia diversa. I raggi cosmici di bassa energia sono contenuti molto bene, invece quelli di energia più elevata presentano una tendenza sempre più spiccata a sfuggire dalla galassia man mano che ci si avvicina all'energia critica; infine, quelli di energia superiore all'energia critica non sono assolutamente soggetti a confinamento.

Non c'è contraddizione tra l'idea di confinamento e la teoria che prevede una origine a due componenti per i raggi cosmici. Secondo questa teoria i raggi cosmici al di sotto di una certa energia sarebbero essenzialmente di origine galattica, mentre quelli al di sopra dell'energia critica sarebbero per lo più extragalattici. Molti teorici dei raggi cosmici furono attratti dal modello a due componenti quando gli esperimenti sugli sciami atmosferici mostrarono che lo spettro dei raggi cosmici non diventava più ripido in prossimità dell'energia critica, come ci si attendeva, ma, al contrario, diventava più piatto. Altri teorici, però, continuarono a difendere l'ipotesi che tutti i raggi cosmici siano prodotti all'interno della nostra galassia. Essi fecero notare che ci sono vari modi per compensare la perdita di raggi cosmici di alta energia nello spazio intergalattico. La nostra conoscenza sugli acceleratori naturali che generano la radiazione cosmica è scarsa; forse lo spettro dei raggi cosmici diventa più piatto perché l'efficienza degli acceleratori cresce notevolmente al di sopra di una certa energia, così da compensare più che ampiamente qualsiasi incremento che possa essere presente nelle perdite dalla Galassia. Pertanto la forma dello spettro di energia dei raggi cosmici non è determinante per decidere tra la teoria dell'origine galattica e quella dell'origine a due componenti; si deve cercare qualche altra prova sperimentale per poter operare la scelta. Le misure della direzione di provenienza degli sciami atmosferici forniscono il tipo di dati che ci occorre.

L'ipotesi del confinamento spiega perché i raggi cosmici di energia bassa o relativamente poco elevata sono quasi isotropi, provengono cioè da tutte le direzioni dello spazio quasi con la stessa intensità. Al di sopra dell'energia critica, però, i raggi cosmici non sono affatto confinati e non possono subire grandi deviazioni per opera del campo magnetico galattico; pertanto le loro direzioni di arrivo devono indicare la localizzazione delle loro sorgenti. Se anche i raggi cosmici di altissima energia sono di origine galattica, allora la maggior parte di essi dovrebbe arrivare dal piano della Via Lattea, mentre solo una frazione molto piccola di essi proverrebbe da direzioni più o meno perpendicolari a tale piano. Se invece i raggi cosmici hanno origine in galassie esterne attive, si dovranno poter individuare alla fine se non altro le più importanti tra quelle galassie.

È interessante pensare a che cosa si potrà scoprire quando si saranno accumulati più dati, soprattutto a energie superiori a 6×10^{19} elettronvolt. Sembra praticamente certo che lo spettro di energia presenti qualche sorta di struttura che può essere posta in relazione con l'effetto di Greisen-Zatsepin, fornendo così un

Le direzioni di arrivo dei 16 sciami più energetici registrati dalla schiera di Haverah Park a partire dal 1963 forniscono l'unica prova sperimentale veramente convincente a favore dell'origine extragalattica dei raggi cosmici più energetici. Ogni sciame ha un'energia di 5×10^{19} elettronvolt. Le singole direzioni sulla sfera celeste sono rappresentate su una proiezione equivalente di Aitoff. La parte di cielo che sta sopra la linea che attraversa il centro della proiezione passa sopra alla schiera ed è scandagliata dalla schiera stessa una volta ogni 24 ore siderali. La croce sopra il centro indica la posizione del polo nord galattico, che si trova in una direzione a 60 gradi circa dalla direzione dell'asse terrestre. Sono indicati anche i paralleli corrispondenti a -30, 0 e $+30$ gradi di latitudine galattica. Se i 16 raggi cosmici energetici fossero stati originati all'interno della Galassia, pochissimi sarebbero arrivati da direzioni comprese entro 30 gradi dal polo nord galattico.

La forma dell'impulso prodotto in un rivelatore presso l'asse dello sciame riflette lo sviluppo longitudinale di uno sciame atmosferico. Il profilo longitudinale di uno sciame può essere misurato rivelando la radiazione emessa al passaggio dello sciame attraverso l'atmosfera; tale profilo fornisce molte informazioni sulla particella che ha prodotto lo sciame. Il principio è illustrato facendo tre ipotesi semplificative: che la regione da cui sono emesse le particelle rivelate abbia dimensioni trascurabili; che tale regione si muova lungo l'asse dello sciame alla velocità della luce, che la radiazione rivelata segua traiettorie rettilinee muovendosi anch'essa alla velocità della luce. L'istogramma al di sopra del livello del terreno è il profilo di emissione, che rappresenta i contributi al segnale luminoso provenienti da intervalli successivi lungo l'asse dello sciame. L'istogramma inferiore è il corrispondente profilo dei tempi di arrivo. Poiché tutte le particelle hanno la stessa velocità, gli intervalli temporali sono proporzionali alle distanze. Il metodo fa uso di rivelatori di muoni e di rivelatori che ricevono la luce Čerenkov atmosferica.

mezzo per verificare l'accuratezza delle misure odierne di energia in questo intervallo, basate sulla analisi dettagliata della struttura degli sciami atmosferici. L'effetto di Greisen-Zatsepin potebbe anche fornire un mezzo indipendente per valutare distanze astronomiche comprese tra 30 milioni e un miliardo di anni luce.

Per essere così ottimista ho dovuto, però, supporre che si possa presto risolvere un grosso problema, di cui non ho ancora parlato. Nei miei commenti sull'interpretazione dei dati esistenti, soprattutto dei dati relativi alla direzione di arrivo dei raggi cosmici, ho utilizzato implicitamente una nota formula per calcolare la curvatura della traiettoria di una particella carica, che si muove perpendicolarmente rispetto a un campo magnetico. Se la particella è altamente relativistica, ovvero se si muove con una velocità molto vicina a quella della luce, il raggio di curvatura è proporzionale al rapporto tra la sua energia totale e l'intensità del campo. Nel caso dei raggi cosmici, le misure effettuate sugli sciami atmosferici ci danno l'energia totale della particella e i dati astronomici ci forniscono una stima dell'intensità del campo magnetico. Il raggio di curvatura è, però, anche inversamente proporzionale alla carica elettrica della particella in movimento, che, nel caso dei raggi cosmici di bassa energia, può variare da 1 (per protoni, ovvero nuclei di idrogeno) a 26 (per nuclei di ferro).

Pertanto, se l'energia critica per il confinamento dei protoni cosmici entro la Galassia è, per un certo modello, di $2,6\times10^{19}$ elettronvolt, allora l'energia critica per il confinamento dei nuclei di ferro presenti nei raggi cosmici sarà, secondo lo stesso modello, 26 volte minore, cioè solo 10^{18} elettronvolt. Segue da ciò che la situazione è molto diversa se i raggi cosmici di altissima energia sono protoni o nuclei di ferro o presentano una qualsiasi altra composizione. La composizione delle particelle che danno inizio agli sciami estesi atmosferici costituisce uno dei problemi insoluti più importanti nella ricerca sui raggi cosmici dotati di altissima energia.

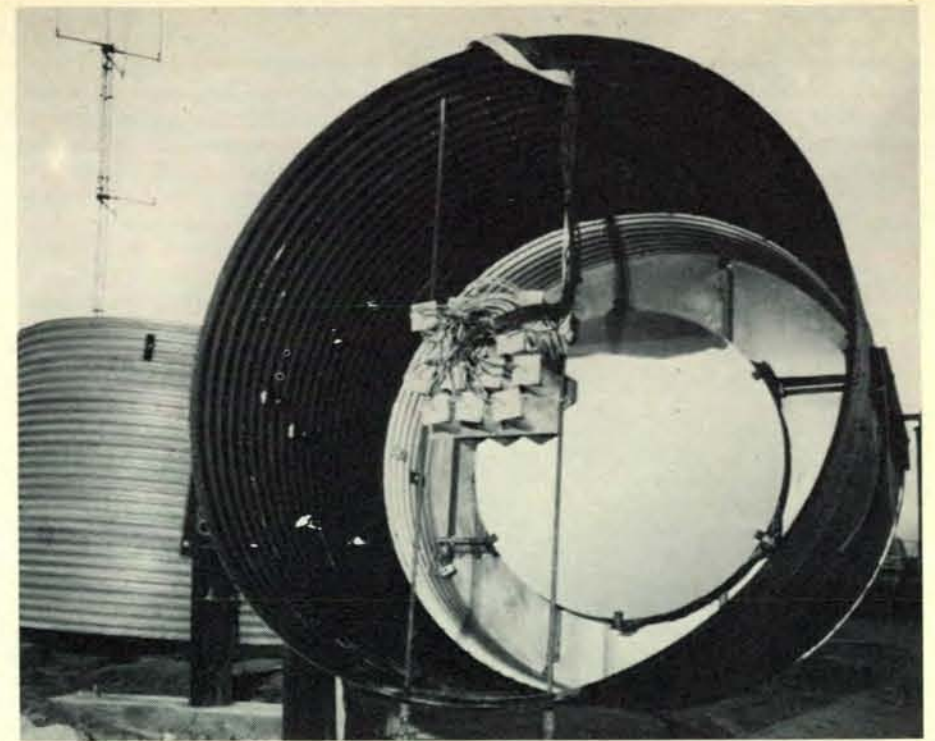
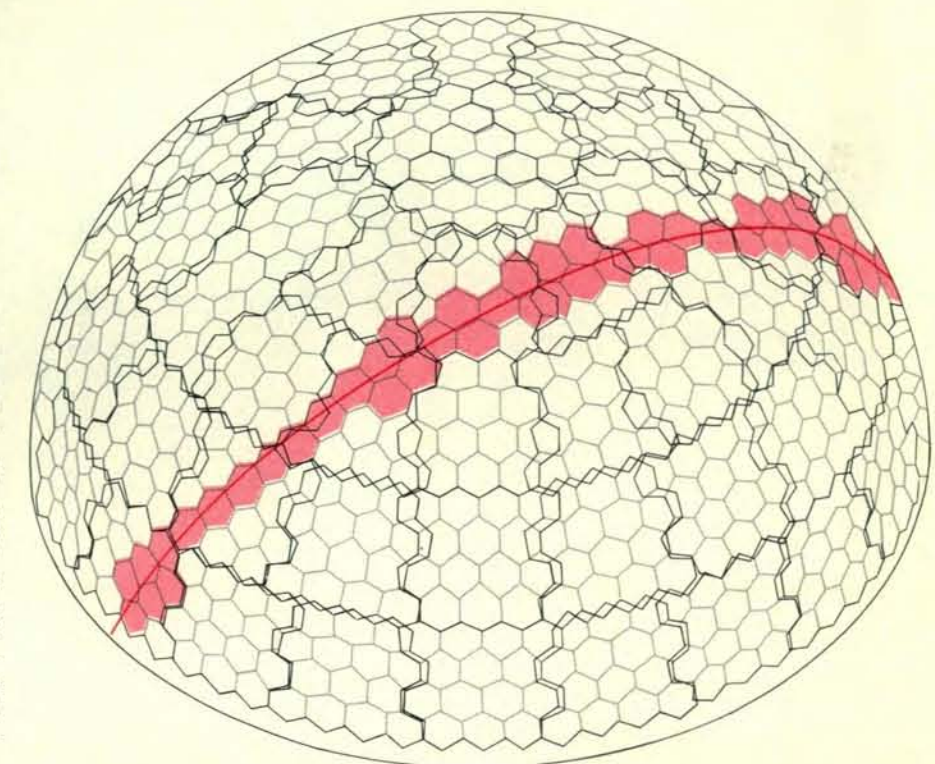
Come si può stabilire quale particella primaria genera gli sciami atmosferici osservati? Le particelle di energia minore che danno luogo a piccoli sciami atmosferici sono tutte indubbiamente nuclei atomici. Questo è sicuro perché i raggi cosmici primari di bassa energia sono stati studiati piuttosto dettagliatamente con rivelatori portati al di sopra dell'atmosfera terrestre. Studi di questo tipo non sono realizzabili a energie maggiori di 10^{15} elettronvolt; la struttura degli sciami atmosferici ci dice però che i raggi cosmici che generano sciami atmosferici di grandi dimensioni sono anch'essi, almeno per la maggior parte, nuclei atomici. Sebbene sia possibile che alcuni raggi cosmici di altissima energia siano particelle subatomiche di massa inferiore a quella dei nuclei, non ci sono dati sperimentali a dimostrare che questa possibilità si verifichi effettivamente.

I metodi sviluppati allo scopo di determinare il contributo dei nuclei atomici dei diversi elementi alla formazione di sciami atmosferici estesi si basano su varie tecniche per valutare due proprietà di uno sciame: la frazione di muoni presenti nello sciame in un istante specifico dello sviluppo e la forma del profilo longitudinale dello sciame (la curva che descrive la crescita e la diminuzione finale del numero di particelle nello sciame). Ci sono buoni motivi per credere che queste proprietà varino in un modo caratteristico, che dipende dal fattore di Lorentz della particella primaria che ha dato inizio allo sciame. Il fattore di Lorentz di una particella in movimento è dato dal rapporto tra la sua energia totale e la sua energia di massa. Al crescere del fattore di Lorentz della particella primaria diminuisce la frazione di muoni presente nello sciame e si allarga il profilo dello sciame stesso.

La frazione di muoni presente in uno sciame può essere valutata usando una schiera di rivelatori del tipo di Williams--Bassi, che misura il numero di particelle di tutte le specie, associata a un rivelatore di particelle veloci schermato con cemento, piombo o terra in quantità sufficiente per assorbire tutti gli elettroni e i fotoni, lasciando passare solo i muoni. Recentemente sono stati sviluppati metodi diretti per registrare profili di singoli sciami: ne parlerò in seguito. In passato si poteva solo cercare di ricavare differenze nei profili di sciami di uguale energia da differenze riscontrate in proprietà correlate col profilo stesso.

Se si esaminano queste proprietà per gli sciami di tutte le dimensioni finora osservate, si può dire che il fattore di Lorentz cresce circa di pari passo con l'energia totale dello sciame. Pertanto si può essere certi che la massa delle particelle primarie che generano sciami di dimensioni diverse non cambia radicalmente.

La teoria che connette le due proprietà al fattore di Lorentz della particella primaria non è però esatta, così come non lo sono le nostre misure. Coi metodi odierni non si riesce a distinguere tra nuclei diversi ma ugualmente energetici, nemmeno tra un protone e un nucleo di ferro. Il problema è serio. Anche se si conosce l'energia di un raggio cosmico e la direzione da cui è entrato nell'atmosfera, come si fa a risalire attraverso il mezzo interstellare, ritracciandone la traiettoria curvata dai campi magnetici, se si conosce la carica solo a meno di un fattore 26? Non va dimenticato che è possibile che i raggi cosmici rimangano indefinitamente l'unico mezzo per studiare le interazioni nucleari a energie maggiori di 10^{15} elettronvolt. Come si può interpretare ciò che accade in collisioni tra particelle dotate di simili energie senza sapere da quali particelle è composto il fascio incidente?

Si sta cercando ora con nuovi esperimenti di misurare direttamente il profilo degli sciami man mano che si sviluppano nell'atmosfera. Si tenga presente il modello di un piccolo disco che scorre lungo un filo. Le particelle all'interno del

Un elemento del nuovo rivelatore «occhio di mosca» per sciami atmosferici, in costruzione a Little Granite Mountain nello Utah, consiste in uno specchio del diametro di 1,5 metri. Tutto il rivelatore sarà costituito da 67 specchi come questo, ciascuno dotato, sul piano focale, di 12 o 14 fotomoltiplicatori. (Questo ne ha 14.) Ogni tubo fotomoltiplicatore rivelerà e registrerà la debole luce emessa nell'atmosfera dagli sciami prodotti dai raggi cosmici. Al completamento, i 67 specchi saranno disposti in modo da formare un «occhio» composto con 880 tubi fotomoltiplicatori.

Il rivelatore occhio di mosca, una volta completato, scandaglierà il cielo alla ricerca di sciami atmosferici da raggi cosmici. Ognuno dei poligoni leggermente sovrapposti, tracciati in nero, rappresenta il campo di vista di uno dei 67 specchi. Ogni poligono è formato da 12 o 14 esagoni, ciascuno dei quali rappresenta i sei gradi del campo di vista di un tubo fotomoltiplicatore. Uno sciame che passi sull'occhio di mosca emetterà una debole luce, attivando i fotomoltiplicatori di cui attraverserà il campo di vista *(in colore)*. Ogni fotomoltiplicatore avrà un canale elettronico per registrare tempi e intensità dei segnali, da cui un elaboratore ricostruirà il profilo dello sciame.

disco generano luce. Pertanto, il disco che costituisce lo sciame atmosferico è come un globo di fuoco, la cui luminosità cresce via via che cresce il numero di particelle, raggiunge la massima brillanza quando lo sciame raggiunge il massimo sviluppo, poi incomincia a indebolirsi. I metodi più recenti per misurare profili di sciami consistono fondamentalmente nel «fotografare» di notte il globo di fuoco in movimento utilizzando la luce che emette. La registrazione dello sciame atmosferico fornita da questi metodi è equivalente a una successione di esposizioni fotografiche.

La luce proveniente dal globo di fuoco è costituita da due componenti dovute a due meccanismi diversi con cui le particelle veloci possono generare luce. Una componente, che in questo caso è nettamente predominante, è emessa in avanti in uno stretto fascio, come la luce proiettata dal faro anteriore di una locomotiva. La componente meno importante è emessa in ogni direzione con la stessa intensità, come la luce proveniente da un meteorite. Il fascio luminoso emesso in avanti è prodotto per effetto Čerenkov ed è indicato come luce Čerenkov atmosferica. La luce emessa isotropicamente è prodotta da fluorescenza. Alcuni nuovi esperimenti sono stati progettati per misurare la luce Čerenkov atmosferica prodotta dagli sciami; altri per valutare la luce emessa per fluorescenza.

La tecnica per fotografare gli sciami atmosferici per mezzo della luce Čerenkov è già stata sviluppata ed è in uso presso due schiere giganti, quella in Inghilterra e quella in Siberia. L'angolo di inquadratura migliore per ottenere esposizioni intense si ha quando lo sciame è proprio diretto verso l'apparecchio fotografico. Il globo di fuoco si muove, però, praticamente con la stessa velocità della luce che emette, così che l'angolo tra l'asse dello sciame e l'asse ottico della macchina da presa non può essere troppo piccolo, altrimenti sarebbe impossibile risolvere il profilo dello sciame. Il miglior compromesso tra la necessità di avere un segnale intenso e quella di avere un buon potere risolutivo si ha con sciami che cadono a varie centinaia di metri dalla macchina da presa. Lo strumento che raccoglie la luce in un esperimento di questo tipo assomiglia poco a una normale macchina da presa. Non ha componenti ottiche. Le esposizioni successive risultano sovrapposte dal punto di vista spaziale; sono invece risolte temporalmente.

Dopo che sono state calcolate le traiettorie degli sciami nel solito modo usando i dati forniti dalla schiera gigante, si può derivare il profilo spaziale apparente da quello temporale utilizzando un semplice metodo grafico. La deduzione del profilo reale è, però, molto più difficile, poiché richiede l'uso di un modello teorico dettagliato delle cascate di nucleoni, con conseguente incertezza sui risultati.

La tecnica per fotografare gli sciami atmosferici per mezzo della luce emessa per fluorescenza atmosferica è ora in via di sviluppo negli USA e in Giappone. Un vantaggio di questo metodo sta nel fatto che in tal modo gli sciami atmosferici sono fotografati di fianco, cioè dall'angolo ideale per avere una buona risoluzione spaziale. Un altro vantaggio sta nel fatto che con questo metodo dovrebbero essere rivelabili sciami atmosferici anche a distanze notevoli, forse fino a decine di chilometri. La macchina da presa necessaria per registrare la luce emessa per fluorescenza deve essere molto più complessa di quella che si usa per registrare la luce Čerenkov; d'altra parte le traiettorie degli sciami possono essere ricavate da un'analisi delle fotografie degli sciami. Pertanto, lo strumento capace di registrare la fluorescenza atmosferica è autosufficiente, poiché non richiede dati ulteriori da una schiera di rivelatori di tipo Williams-Bassi. Infine, il vantaggio forse principale: sono necessarie solo piccole correzioni per derivare il profilo reale di uno sciame dalle immagini della luce emessa per fluorescenza.

I grandi vantaggi potenziali connessi alla rivelazione degli sciami per mezzo della fluorescenza atmosferica erano noti almeno fin dal 1962, ma si conoscevano anche le difficoltà, prima tra tutte la bassissima intensità di tale luce. Per ottenere anche risultati minori, occorrerebbero molti grandi collettori di luce dotati di notevole efficienza e posti in un luogo ideale, con aria tersa, clima favorevole e una quantità minima di luce di fondo. Tra il 1965 e il 1974 si fecero numerosi tentativi per costruire un'apparecchiatura efficiente, ma fallirono. Due anni fa un gruppo di ricercatori dell'Università dello Utah guidato da Haven E. Bergeson portò a Volcano Ranch tre nuovi prototipi di collettori di luce di elevata sensibilità con cui fece prove soddisfacenti operando in connessione con la schiera di Volcano Ranch. Si sta ora costruendo uno strumento completo, composto da 67 collettori di luce, il campo di vista dei quali è suddiviso tra 880 fotomoltiplicatori. Quando i collettori saranno terminati, saranno montati sulla Little Granite Mountain a Dugway Priving Grounds, 160 chilometri a ovest di Salt Lake City.

Lo strumento dello Utah è stato chiamato «occhio di mosca». Ogni fotomoltiplicatore controlla il proprio piccolo campo di vista esagonale, analogamente a quanto fa ogni elemento dell'occhio composto di una mosca. Per comodità di funzionamento i collettori sono montati separatamente, ciascuno nel proprio alloggiamento rotante. Dopo il tramonto, se il tempo è buono, i collettori si dispongono automaticamente in posizione, così che insieme guardano tutta la volta celeste. I segnali provenienti dai tubi fotomoltiplicatori saranno subito analizzati da un piccolo elaboratore elettronico in linea. Il segnale caratteristico di uno sciame atmosferico è un globo di fuoco che si muove alla velocità della luce su una traiettoria rettilinea. Quando l'elaboratore riconoscerà un evento dotato di questo aspetto caratteristico, comanderà una ripresa ad alta velocità, in codice, che sarà analizzata successivamente. Con questa registrazione si potrà ricostruire la traccia dello sciame nelle tre dimensioni spaziali. Questa ricostruzione fornirà subito l'energia e la direzione di arrivo del raggio cosmico primario che ha originato lo sciame.

Se il comportamento dell'occhio di mosca sarà pari alle aspettative, esso accumulerà dati sui raggi cosmici più energetici con una rapidità 10 volte superiore a quella di tutte le schiere giganti finora esistenti messe insieme. I profili degli sciami misurati direttamente che esso fornirà dovrebbero giocare un ruolo importante nella risoluzione del problema della composizione, superando così uno degli ultimi ostacoli che impediscono di scoprire l'origine dei raggi cosmici ad altissima energia.

La traiettoria di uno sciame atmosferico potrà essere calcolata coi dati registrati dall'occhio di mosca. Dato che la luce emessa dagli sciami atmosferici è tanto debole, sarà rivelabile solo di notte. I globi di fuoco degli sciami atmosferici si muovono in linea retta praticamente alla velocità della luce. Le quattro linee che si dipartono dalla traiettoria dello sciame atmosferico *(in colore)* rappresentano percorsi seguiti dalla luce emessa dallo sciame per raggiungere due coppie diverse di fotomoltiplicatori del rivelatore. Il ritardo temporale tra i segnali provenienti da fotomoltiplicatori adiacenti varierà in funzione della relazione geometrica tra il rivelatore e la traiettoria dello sciame.

Immagini notturne della Terra riprese dallo spazio

Le immagini registrate a mezzanotte dai satelliti meteorologici mostrano un aspetto inedito della Terra: le luci più intense sul lato oscuro del pianeta sono gigantesche fiamme di gas di scarico

di Thomas A. Croft

Quando la Terra viene osservata dallo spazio durante il giorno, la scena è dominata dalle nuvole e dalle caratteristiche geografiche. Le prove della presenza umana, quando sono visibili, costituiscono una modificazione secondaria del quadro d'insieme. Di notte, invece, e specialmente a mezzanotte di una notte senza Luna, l'immagine scura della Terra scintilla per le luci brillanti delle creazioni umane. Oltre ai mutevoli chiarori di fenomeni naturali quali i fulmini, gli incendi di foreste, le aurore boreali, il lato notturno della Terra si ammanta dei complicati ricami delle luci delle città e delle zone, più sparse, dei fuochi agricoli. Ma fra le opere dell'uomo quelle di gran lunga più luminose sono le fiamme di gas che sono associate ad alcuni dei più grandi campi petroliferi del mondo.

Questo aspetto inedito della Terra è divenuto evidente solo recentemente, come risultato imprevisto del Defence Meteorological Satellite Program dell'US Air Force. Lo scopo principale della navicella messa in orbita nel corso di questo programma è di fornire immagini sulle quali basare le previsioni a breve scadenza della nuvolosità. Dapprima le immagini, trasmesse elettronicamente sulla Terra, venivano normalmente gettate via dopo essere state usate un giorno, ma presto si riconobbe lo spreco di potenziali dati scientifici e vennero prese delle misure per conservare le migliori immagini di aurore boreali. Infine tali immagini divennero accessibili al pubblico nel 1973. Si stabilì la costituzione di un archivio dei negativi originali nell'Università del Wisconsin e la distribuzione su microfilm delle immagini ai ricercatori interessati, attraverso l'organizzazione del Dipartimento del commercio. Quasi tutte le fotografie di questo articolo sono state ottenute dall'archivio del Wisconsin.

Caratteristiche del sistema

Il satellite meteorologico dell'Air Force non è l'unico sistema in grado di fornire immagini notturne della Terra vista dallo spazio. Anche i tre veicoli spaziali Landsat, lanciati nel 1972, 1975 e 1978 dalla NASA, potrebbero funzionare di notte, per quanto non siano stati progettati per questo scopo e raramente siano stati usati per fare fotografie notturne. Il sistema Landsat, rispetto ai satelliti meteorologici dell'Air Force, ha il vantaggio di fornire immagini a quattro colori e con una risoluzione a terra migliore, ma l'altro sistema è molto più sensibile a deboli sorgenti luminose. Per lo studio delle sorgenti luminose più intense i due sistemi sono complementari: il satellite dell'Air Force è adattissimo a compiere una ricognizione a grande scala di tutta la Terra, mentre il sistema Landsat può fornire fotografie a colori ad alta risoluzione di certe aree particolari che vengono selezionate dalla ricognizione generale.

Ambedue questi programmi hanno veicoli spaziali in orbite quasi circolari, a quote di circa 800 o 900 chilometri e le orbite sono studiate in modo da passare a circa nove gradi dai poli della Terra e da far compiere al piano orbitale una precessione attorno alla Terra alla stessa velocità con cui la Terra si muove attorno al Sole. Orbite di questo tipo, sincronizzate col Sole, offrono un vantaggio importante per le missioni fotografiche: al di sotto della navicella l'ora del giorno (o della notte) non varia al passare delle stagioni, e quindi si minimizzano le variazioni di illuminazione. L'effetto pratico di questa situazione è che le immagini vengono registrate sotto forma di strisce successive, in direzione approssimativamente nord--sud, in cui ogni striscia rappresenta un'orbita del percorso del satellite. L'ora del giorno in ogni zona fotografata non cambia da striscia a striscia. La sovrapposizione fra strisce adiacenti aumenta alle alte latitudini: a 90 gradi di latitudine nord e sud i poli compaiono in ogni striscia; essi vengono fotografati alternativamente ogni 102 minuti.

Una delle cose più sorprendenti di molte immagini ottenute da questo sistema è la sorgente luminosa: la luce della Luna a mezzanotte. Alla luce del giorno, nell'altra metà dell'orbita, il satellite registra della Terra un quadro simile, ma più dettagliato. Di conseguenza il satellite, ogni 12 ore circa, fornisce fotografie della nuvolosità su tutta la Terra, nei periodi in cui la Luna illumina il lato buio della Terra (e cioè circa la metà del tempo). Su tutti questi satelliti vi sono diversi altri tipi di sensori. Un apparecchio fa contemporaneamente fotografie all'infrarosso, in base alle radiazioni termiche della Terra e delle nuvole; le pellicole così ottenute mostrano le temperature relative delle diverse nubi e permettono di mappare la nuvolosità anche senza Luna.

Nelle fotografie fatte alla luce della Luna piena, quando le apparecchiature fotografiche del satellite funzionano a sensibilità ridotta, si vedono molti fuochi e luci prodotti dall'uomo, ma solo le fiammate di gas sono abbastanza luminose da imporsi chiaramente *(si veda l'illustrazione alle pagine 46 e 47)*. Invece, al calare della luce della Luna, le fotografie subiscono una trasformazione assai simile a quella che si verifica in cielo di sera, quando il sole tramonta e compaiono le stelle, salvo che qui compaiono le città e i fuochi! Consideriamo ad esempio la fotografia della pagina a fianco. A nord, nella luce incerta della Luna, compare chiaramente lo stivale dell'Italia, e si possono localizzare facilmente molte delle città principali d'Europa. I punti più luminosi di questa striscia sono fiammate di gas che coincidono coi campi petroliferi conosciuti dell'Algeria, della Libia e della Nigeria. La zona abbastanza uniforme di lucine sparse per l'Africa a sud del Sahara è data dai fuochi appiccati deliberatamente per liberare terreni dedicati alla pastorizia o all'agricoltura. Si vedono incendi simili in molte altre parti del mondo.

In questa fotografia si vedono molte delle imperfezioni del sistema del satellite meteorologico dell'Air Force. I dettagli presso i bordi sono parecchio annebbiati, a causa della bassa quota del satellite, che è circa un terzo della larghezza dell'immagine. Oggi i satelliti meteorologici dell'Air Force possiedono apparecchiature di scansione modificate, con caratteristiche in grado di migliorare la risoluzione ai bordi di queste immagini.

Il sistema ottico che produce queste fotografie funziona nel modo seguente. L'immagine ha origine in un telescopio ottico che esplora la superficie terrestre in direzione perpendicolare alla direzione del moto del satellite, con risoluzione angolare costante. (Nelle zone non polari l'esplorazione è grosso modo est-ovest.) Nel piano focale del telescopio c'è un elemento fotosensibile, che assomiglia a un esposimetro fotografico, salvo per il fatto che abbraccia un campo di soli due decimi di grado. I bordi della scena si trovano lontani dal satellite e vengono osservati obliquamente; quindi l'angolo abbracciato dal telescopio copre un'area superficiale circa sei o sette volte maggiore dell'area vista al nadir (cioè quando la fotocamera punta diritta verso il basso). I dati di questo sistema vengono inviati a terra sotto forma di una sequenza di numeri binari, che rappresentano le variazioni di luminosità registrate dal sensore durante la sua esplorazione. Successivamente, sulla Terra, i numeri vengono riconvertiti in punti luminosi, e le immagini vengono ricostruite su pellicola in forma convenzionale, come una mappa. Nel corso di questo processo si fanno delle correzioni approssimate delle distorsioni, che possono essere causate da svariati motivi. (L'orbita non è esattamente circolare, la Terra non è esattamente sferica, il satellite non è sempre alla stessa quota.)

Alcune delle maggiori fiammate di gas sono così luminose che diventano visibili le riflessioni interne al corpo del telescopio e questo effetto spiega gli anelli che in certe immagini circondano le fiammate di gas. Effetti simili compaiono anche nei telescopi astronomici, che sono strumenti progettati per rivelare corpi piccolissimi, ma molto luminosi rispetto allo sfondo.

Paragonando queste immagini con quelle corrispondenti fatte all'infrarosso, che hanno esattamente le stesse dimensioni e quindi sono facilmente sovrapponibili su un tavolo luminoso, si vede che le luci delle città penetrano le nubi sottili attenuandosi di poco; le nuvole non molto dense diffondono l'immagine, mentre quelle molto pesanti possono bloccare la luce completamente.

Fotografia notturna da satellite, fatta recentemente nel corso del Defense Meteorological Satellite Program dell'US Air Force, che mostra tre sorgenti luminose principali, associate ad attività umane. I numerosi punti luminosi disposti nel terzo superiore della foto sono luci di città europee. Le luci più grandi isolate verso il centro e in basso sono dovute a fiammate di gas dei campi petroliferi dell'Algeria, della Libia e della Nigeria. (Gli anelli attorno alle fiammate più intense sono provocati da riflessioni interne al corpo del telescopio.) La fascia uniforme di luci più piccole sparse attraverso l'Africa a sud del Sahara è dovuta a incendi appiccati a scopi agricoli e di pastorizia.

Le fiammate di gas

Le immagini notturne da satellite mostrano fiammate luminose in molte parti del mondo, ma in prossimità del Golfo Persico si ha di gran lunga la massima concentrazione; anche alla luce della Luna piena questa zona brilla di numerose grandi fiammate *(si veda l'illustrazione a pagina 48 in basso)*. L'incendio di gas di scarico nei campi petroliferi del Golfo Persico è da decenni uno spettacolo eccezionale in questa parte del mondo, descritto già nel 1933 da aviatori che sorvolavano questa zona. L'uso di liberare in aria i gas indesiderati dei pozzi petroliferi senza bruciarli risale ancora più addietro: più di 100 anni fa un articolo di «Scientific American» descriveva lo scarico di gas dei primi pozzi petroliferi della Pennsylvania; l'autore stimava che un pozzo liberasse circa 30 000 metri cubi di gas all'ora e che lo facesse «da anni».

Siccome le fiammate di gas sono le caratteristiche più notevoli di queste immagini notturne, esse meritano una maggiore attenzione. Il sistema fotografico del satellite meteorologico dell'Air Force è progettato per misurare soltanto la luminosità media di ciascun elemento della fotografia, che viene detto *pixel*, e che misura circa tre chilometri di lato. (Le dimensioni esatte variano da luogo a luogo sull'immagine e da sensore a sensore.) Di conseguenza il contributo di una singola fiamma alla luminosità media deve dipendere non solo dalla sua luminosità intrinseca, ma anche dalla percentuale di area del pixel che è occupata dall'immagine della fiamma. In altri termini, per risultare sulle immagini così chiaramente, le fiammate devono essere o molto luminose o molto grandi; e in effetti esse sono ambedue le cose. La vasta area coperta dalle fiammate è una conseguenza della pratica di far espandere il fuoco lateralmente per assicurare la combustione completa e di solito questo si ottiene liberando il gas da una serie di ciminiere disposte su una fila lunga 50 metri o più *(si veda l'illustrazione in alto a pagina 49)*.

Nel tentativo di analizzare con più dettaglio due delle maggiori fiammate di gas dell'Algeria, che si vedono nell'illustrazione alla pagina precedente, mi sono procurato le fotografie fatte dal Landsat, di giorno, della stessa zona *(si veda l'illustrazione in basso a pagina 49)*. In parecchie di queste fotografie si vedono chiaramente due fiammate, a una distanza nella direzione nord-sud di circa 11 chilometri; nei termini del sistema fotografico del satellite dell'Air Force esse sono a una distanza di tre pixel.

Studiando con la lente d'ingrandimento delle diapositive ingrandite delle fotografie del Landsat, sono riuscito a stabilire che le fiammate rosse hanno un diametro di circa 500 metri, molto più grande di quanto si vede normalmente nelle fotografie scattate al livello del suolo. È facile individuare le fiammate di gas nelle fotografie fatte dai satelliti di giorno, perché il loro fumo si estende per decine di chilometri nella direzione del vento. Le fiammate di gas più moderne sono fornite di bruciatori «senza fumo», per evitare di dare una prova così evidente di inquinamento, e questa pratica è senza dubbio giustificata dal punto di vista delle apparenze; ma dal punto di vista del preoccupante aumento del livello di anidride carbonica nell'atmosfera, a lungo termine l'umanità potrebbe cavarsela meglio se una parte delle particelle carboniose rimanesse incombusta.

Dalle immagini notturne del Landsat si può ricavare un'idea più precisa sull'estensione orizzontale di queste fiammate. A tale scopo una navicella spaziale Landsat venne attivata sull'Algeria la notte del 31 gennaio 1976, su richiesta dell'US Geological Survey. L'illustrazione in alto a pagina 50 è un ingrandimento estremo di una delle fiammate fotografate in questa occasione. La fiammata corrisponde a quella più grande e più a sud che si vede nella foto diurna del Landsat a pagina 49 in basso.

Per capire il significato di questa particolare immagine notturna, bisogna sapere qualcosa di come è stata fatta. Durante il passaggio del satellite Landsat, il suo telescopio a scansione multispettrale guarda verso il basso attraverso uno specchio a 45 gradi che oscilla avanti e indietro per esplorare la scena da sinistra a destra (cioè in una direzione perpendicolare a quella del movimento). Le strutture luminose che cadono sul piano focale del telescopio vengono trasmesse ai sensori elettronici per mezzo di sottili tubicini di vetro, ognuno dei quali ha una sezione che corrisponde come area a un quadrato di 79 metri di lato sulla Terra, al nadir. (Su un bordo dell'immagine, a 93 chilometri dalla proiezione dell'orbita del satellite sulla Terra ogni guida ottica copre una superficie leggermente più grande e quindi il potere risolutivo teorico del sistema di solito viene arrotondato a 80 metri.) I sensori elettronici vengono «letti» ogni 57 metri di esplorazione e quindi i pixel del Landsat sono rettangoli, che al nadir misurano 57 metri per 79. Un grande oggetto luminoso può influire su due o tre letture consecutive del sensore, in parte perché le guide ottiche sono più grandi del pixel di 57 metri e in parte a causa di ritardi nell'elettronica. (Alcuni mesi fa i delegati sovietici alle Nazioni Unite hanno affermato che non dovrebbero essere permessi dei pixel più piccoli di 50 metri; una risoluzione maggiore verrebbe considerata una violazione del diritto di una nazione alla sua *privacy*.)

Da ogni pixel il sistema fotografico Landsat ricava quattro letture diverse del sensore, una per ciascuna di quattro bande spettrali, o colori. (Il termine «colore» qui viene usato in senso lato; in effetti due bande spettrali sono una nell'infrarosso e l'altra nell'infrarosso vicino e quindi sono invisibili per l'occhio umano.) Per ogni colore e per ogni pixel il Landsat manda a terra un numero binario in codice, che rappresenta l'intensità della luce. Poiché nel particolare ingrandimento che stiamo considerando ci sono soltanto 35 pixel illuminati, è stato possibile riprodurre singolarmente ogni rettangolino colorato

Panorama al chiaro di Luna di parte della Terra, che si estende dall'Africa nordoccidentale all'Asia sudorientale, composto da parti di sei immagini riprese in sei orbite successive dal satellite meteorologico dell'Air Force. Tutte le fotografie sono state scattate alla mezzanotte locale del 6 febbraio 1974; l'illuminazione è data dalla luce riflessa della Luna. La sovrapposizione fra strisce adiacenti aumenta alle alte latitudini; in questo caso le strisce sono state tagliate in alto e in basso, mentre ogni striscia originale comprende ambedue i poli. Per quanto in un ambiente così intensamente illuminato dalla Luna si vedano anche altre luci, solo le fiammate di gas emergono chiaramente. L'illustrazione in basso alla pagina seguente mostra un ingrandimento di un gruppo di brillanti fiammate di gas di scarico nell'area del Golfo Persico.

Luci urbane dei paesi che costeggiano la Manica e il Mare del Nord, visibili con chiarezza insolita in questo ingrandimento di una foto da satellite dell'Air Force, nonostante il contrasto sia basso. La ripresa è stata fatta in una notte senza Luna del maggio 1977, quando i sensori operavano alla sensibilità massima. Le isole britanniche sono a sinistra. Le luci al centro verso destra sono in Belgio mentre il punto luminoso presso il margine in basso a destra è Parigi.

Le fiammate di gas si concentrano particolarmente nella zona del Golfo Persico, come viene messo in evidenza in questo ingrandimento di una fotografia alla luce della Luna, fornita dal sistema ottico del satellite dell'Air Force degli Stati Uniti. La zona scura in alto è il Mar Caspio, il punto chiaro sul promontorio è Baku, in URSS; il punto più grande a sud del Mar Caspio è Teheran, in Iran, e il punto vicino ai grandi laghi al centro verso sinistra è Baghdad, in Iraq.

con mezzi fotografici ordinari. Per rendere visibili i dati ottenuti all'infrarosso e all'infrarosso vicino ho adottato un metodo di riproduzione a «colori falsati», spostando ogni livello digitale d'intensità verso la zona dello spettro a lunghezze d'onda più brevi, di circa un quinto di micrometro. Quindi le letture all'infrarosso vengono riprodotte in rosso, l'infrarosso vicino in verde e il rosso in blu. (Nell'immagine originaria il verde era praticamente assente.)

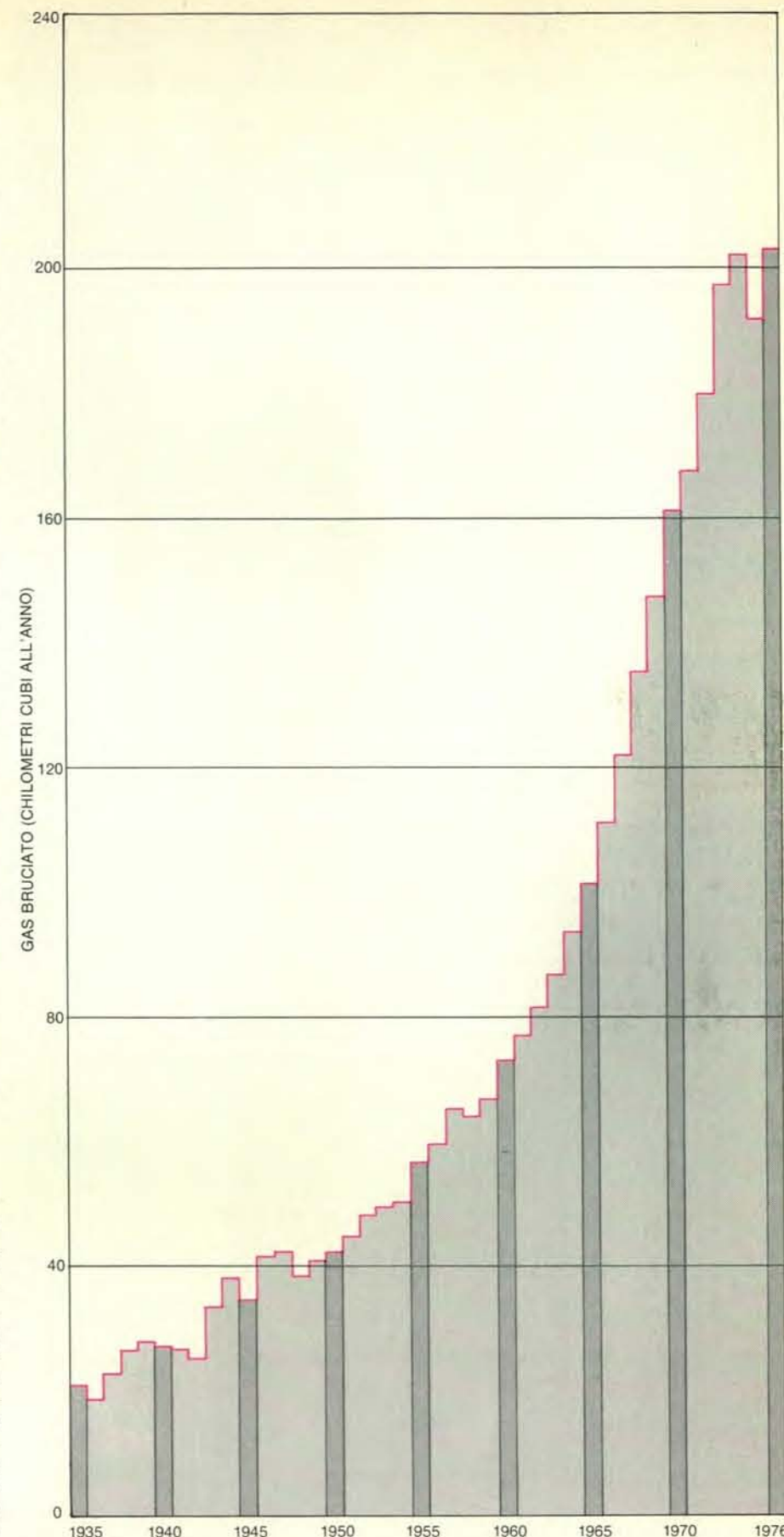
Una normale immagine fotografica del Landsat è costituita da 7 581 600 pixel, ognuno con quattro diverse letture spettrali; ne consegue che un'unica immagine diurna rappresenta più di 30 milioni di numeri trasmessi dallo spazio. Per ricostruire su una pellicola un'immagine a colori di questo tipo non si potrebbe usare nessun metodo fotografico semplice; e in effetti il metodo che si usa per ricavare le fotografie del Landsat non impiega neanche la luce. La pellicola viene invece posta nel vuoto ed esposta al bombardamento diretto di un fascio di elettroni. Questo sistema permette di esporre linee sottilissime di pellicola, e ogni linea corrisponde a una serie di pixel contigui, ottenuti da una esplorazione del telescopio. La composizione dei colori viene poi effettuata meccanicamente sovrapponendo le pellicole in bianco e nero, ognuna delle quali rappresenta una delle bande spettrali originarie. La ricostruzione delle immagini dei satelliti dell'Air Force viene fatta in un modo molto simile, salvo il fatto che si usa un fascio di luce invece che un fascio di elettroni.

Tenendo presenti queste informazioni sul funzionamento del sistema Landsat, torniamo ora a esaminare l'illustrazione in alto a pagina 50 e cerchiamo di visualizzare la fiamma che ha prodotto questa luce. Molti dei pixel ai margini della foto non mostrano che una debole illuminazione, cosa che ci si può aspettare se in quel rettangolo è entrata soltanto una propaggine della fiamma. Per quanto l'immagine totale sia lunga circa 500 metri, si vede che potrebbe essere stata provocata da due fiammate distanti circa 200 metri; d'altra parte il centro oscuro potrebbe essere fumo che oscura in parte la fiamma. (Nell'immagine Landsat fatta di giorno dello stesso incendio le fiamme non sembravano separate.) Probabilmente questo punto potrebbe essere risolto con l'esame di parecchie altre immagini Landsat dello stessso luogo.

Il confronto delle immagini a grande scala dell'Air Force con le carte pubblicate dei principali campi di estrazione di petrolio e di gas del mondo mostra che molte di queste fiammate coincidono con le zone di estrazione di petrolio, ma poche con quelle di estrazione di gas. Ciò non è sorprendente; un pozzo per l'estrazione di gas si può tappare facilmente, se il gas non si può vendere, mentre è inevitabile che un pozzo petrolifero scarichi del gas durante l'estrazione del petrolio. Spesso il petrolio giunge alla superficie sotto forma di un miscuglio schiumoso di vari gas e deve depositare prima di poter essere immesso in un oleodotto. Bisogna decidere cosa fare dei gas associati e, in zone isolate, può essere più economico bruciarli piuttosto che trasportarli fino a un acquirente. I politici che stanno al governo farebbero bene a valutare il significato di queste fotografie, prima di fissare dei controlli sul prezzo del gas naturale per usi domestici. Il contenimento artificiale del prezzo del gas negli Stati Uniti e in altri paesi sviluppati può essere visto come un incoraggiamento verso questi sprechi nei campi petroliferi lontani di paesi meno sviluppati. Presumibilmente i politici dei paesi del mondo a economia controllata si trovano di fronte a un dilemma economico analogo, perché in URSS e in altri paesi comunisti si vedono grandi fiammate di gas associate a giacimenti petroliferi.

Siccome l'umanità sta buttando al vento o bruciando questo gas da più di 100 anni, è naturale chiedersi quanto ne sia stato sprecato in questo modo. È difficile dare una risposta precisa, in quanto le parti più direttamente interessate hanno pochi motivi di essere franche sui dettagli. Nonostante tali difficoltà Ralph M. Rotty, dall'Oak Ridge Institute for Energy Analysis, ha compilato stime annue del volume totale di gas incendiato nel mondo dal 1935 fino al 1976 (*si veda l'illustrazione a destra*). I suoi dati indicano che la quantità di gas incendiato ha avuto un aumento continuo che supera il tasso di crescita di un semplice modello esponenziale, almeno prima dei recenti aumenti del prezzo del petrolio. Sommando queste valutazioni annue e portandole avanti fino al 1978 ammettendo uno scarico costante, e indietro, esponenzialmente, fino alla scoperta del petrolio, avvenuta nel 1859, sono giunto a un volume totale di gas sprecato (alla pressione atmosferica) di 4200 chilometri cubi. Rotty valuta che in anni recenti il gas incendiato comprenda circa il 3 per cento di tutti gli idrocarburi bruciati dall'uomo. Le osservazioni notturne da satellite che ho descritto potrebbero offrire uno strumento indipendente per rideterminare quanto gas venga correntemente incendiato nel mondo.

Lanterne giapponesi

L'immagine in basso della pagina seguente mostra una delle cose più strane che si vedono sulla faccia buia della Terra. A destra c'è il Giappone, che di notte è una delle nazioni più illuminate del mondo, a causa della densità della popolazione e dell'avanzato stato di sviluppo. A sinistra, in mezzo al Mar del Giappone, in un'area che normalmente è nera in una notte così buia, c'è una gran massa vorticosa di luci brillanti. Le luci sembrano disposte in modo ordinato e la loro disposizione cambia di notte in notte. Si sa da molto tempo che questa zona è soggetta a violente tempeste, con uno spiegamento di fulmini straordinario; però, in risposta alla mia domanda, il Servizio meteorologico del Giappone dichiarò che la notte in cui l'immagine era stata registrata nel Mar del Giappone non c'era stata nessuna tempesta di rilievo. Una indagine succes-

Valutazione del volume totale di gas (alla pressione atmosferica) incendiato in tutti i pozzi petroliferi del mondo fra il 1935 e il 1976, compilata da Ralph M. Rotty, dell'Oak Ridge Institute for Energy Analysis. Riportando questi dati all'indietro fino alla scoperta del petrolio, (1859), e avanti fino a oggi, si ha un totale di 4200 chilometri cubi di gas sprecato in poco più di un secolo.

Gigantesca fiammata di gas in Siberia in una immagine notturna da satellite scattata nel gennaio 1975. Questa fiammata, una delle più grandi trovate finora dall'autore nel suo esame delle fotografie dell'Air Force, si trova presso Surgut, nella Siberia settentrionale. I grandi giacimenti petroliferi di questa lontana regione evidentemente sono privi di un oleodotto che porti il gas di scarico in zone in cui se ne potrebbe fare un uso più produttivo. Il percorso della transiberiana è segnato dalle città grandi e piccole distribuite a collana in basso nella foto.

Luci brillanti nel Mar del Giappone appaiono a sinistra in questa fotografia fatta da satellite dell'Air Force in una notte senza Luna del giugno 1975. Le luci coincidono con la posizione della flotta giapponese per la pesca dei calamari in quel momento. Le varie migliaia di pescherecci portano lampade a incandescenza per un totale di milioni di watt, che servono ad attirare i calamari in superficie. Il Giappone stesso è delineato dalle luci urbane delle sue regioni costiere.

siva ebbe una risposta più soddisfacente. Gli scienziati del Ministero della pesca giapponese mi diedero delle carte che mostravano l'ubicazione della flotta per la pesca dei calamari nella data in questione. La coincidenza fra le carte e le immagini del satellite era chiarissima.

Venni informato del fatto che durante la stagione della pesca dei calamari, delle grandi flotte di navi, sia coreane che giapponesi, sono al lavoro nel Mar del Giappone. La sola flotta giapponese è composta da circa 2600 navi, ognuna con stazza da 60 a 100 tonnellate; e ognuna di esse trasporta fino a 50 lampade a incandescenza, della potenza media di 3500 watt per lampada; il che porta a un totale di più di 400 milioni di watt di energia elettrica! Le lampade vengono usate per attirare il calamaro in superficie, dove viene catturato con la rete. Ogni peschereccio porta le sue lampade in una doppia fila, a circa due o tre metri di altezza dal ponte; la metà delle lampade non ha riflettore e l'altra metà ha dei piccoli paralumi. In media soltanto la metà circa della flotta giapponese pesca contemporaneamente e quindi la luce totale emessa potrebbe essere solo dell'ordine di 200 milioni di watt; ma anche così la zona dev'essere veramente illuminata a giorno. Per di più, oltre a queste navi, c'è la flotta coreana, più piccola, sempre per la pesca dei calamari, e un'altra flotta giapponese per la pesca dei lucci in acque meno profonde.

Ci si può fare un'idea più chiara dell'entità di questo wattaggio paragonandolo alla potenza totale mandata in cielo dalla popolazione degli Stati Uniti sotto forma di luce visibile. Questa energia è stata misurata dallo spazio dieci anni fa e si è trovato che è circa 40 milioni di watt - circa un quinto di watt a persona. Tenendo conto della bassa efficienza delle lampade a incandescenza dei pescatori, rispetto alla combinazione di luci usate di notte negli Stati Uniti, credo che la quantità totale di luce vista dal satellite dell'Air Force proveniente dagli Stati Uniti sia solo due o tre volte più grande di quella proveniente dalla flotta per la pesca dei calamari nel Mar del Giappone.

Certamente le luci urbane di ogni nazione, compresi gli Stati Uniti, sono molto più disperse. In effetti la loro distribuzione riflette abbastanza accuratamente la distribuzione della popolazione. Questa corrispondenza si può apprezzare facilmente paragonando il montaggio di tre immagini, riprese da satelliti, degli Stati Uniti di notte, nella pagina a fronte in alto, con la mappa della popolazione degli Stati Uniti, elaborata al calcolatore, riprodotta nella stessa pagina in basso. C'è solo qualche discrepanza; ad esempio a est di Salt Lake City compaiono due luci che non hanno un corrispettivo sulla carta della popolazione.

Incendi agricoli e incendi naturali

In molte zone agricole della Terra vengono bruciati regolarmente vegetali come parte del ciclo stagionale della coltivazione. Ho già menzionato la fascia di incendi

Nelle zone isolate le fiammate di gas vengono fatte espandere su una grande superficie, perché la combustione sia completa. Questa fotografia, scattata da un punto elevato di una torre di perforazione vicina, mostra una di tali fiamme in un campo petrolifero del Kuwait. Questa fotografia è stata fatta da David F. Cupp ed è stata pubblicata per la prima volta sul numero di «National Geographic» del maggio 1969.

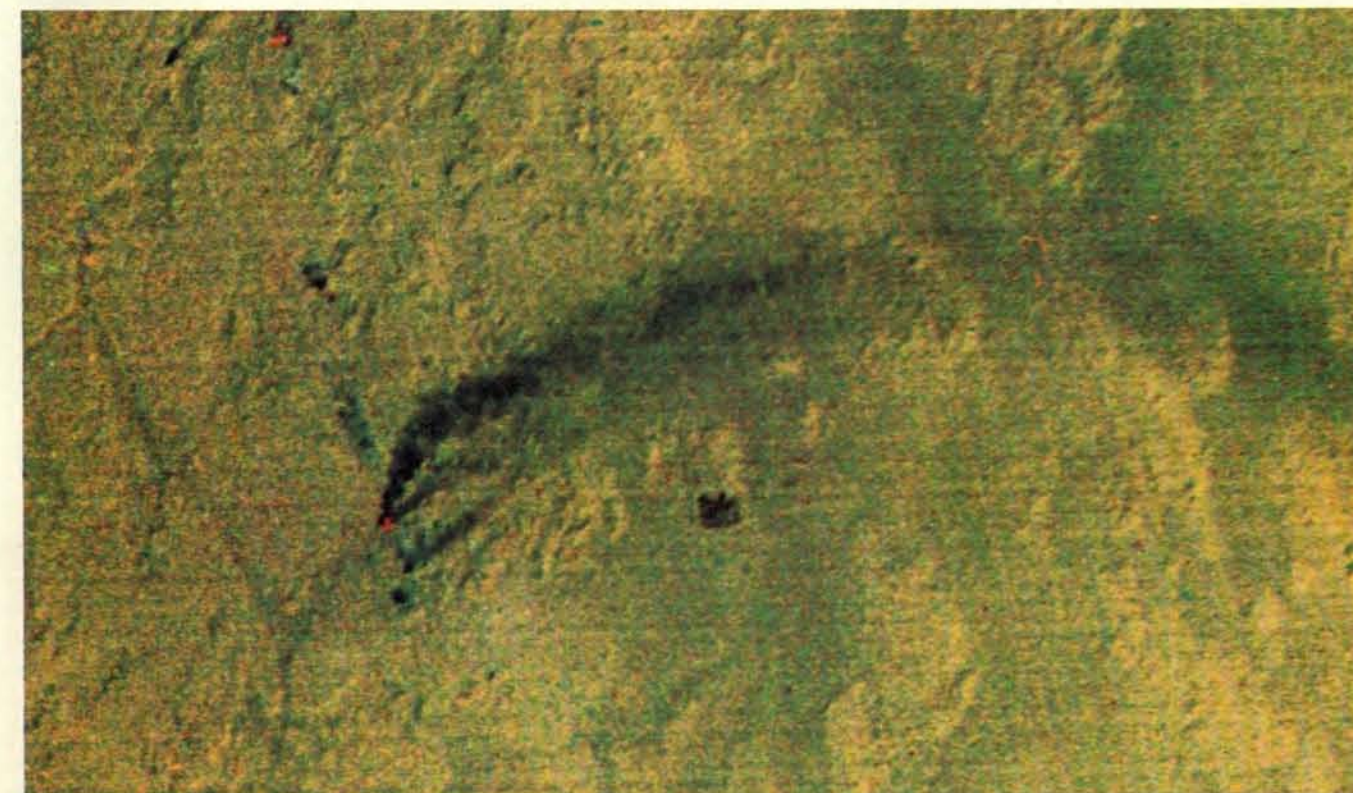
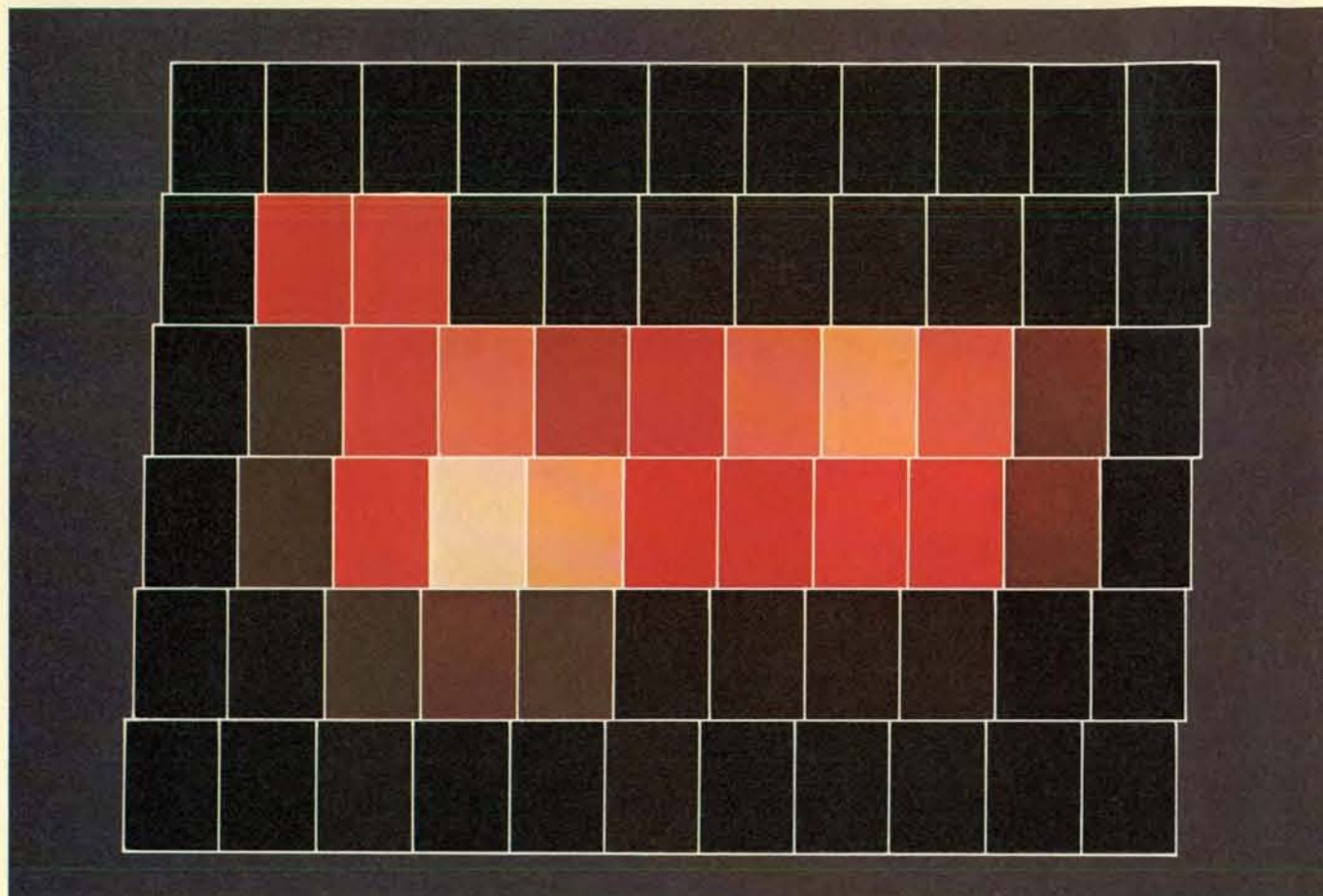
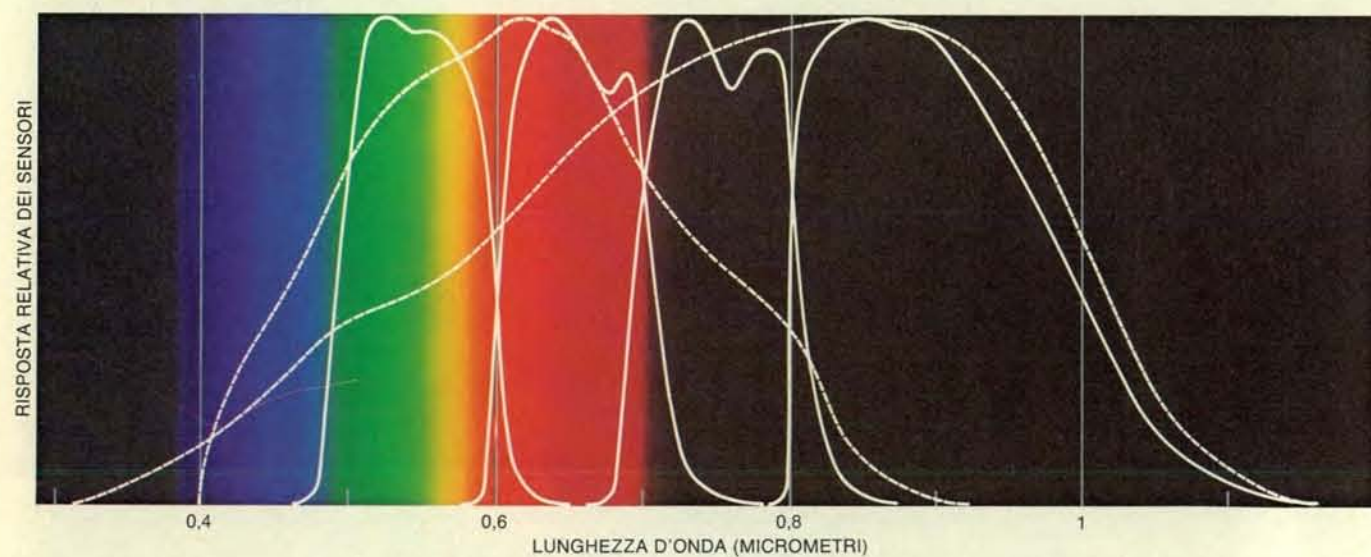
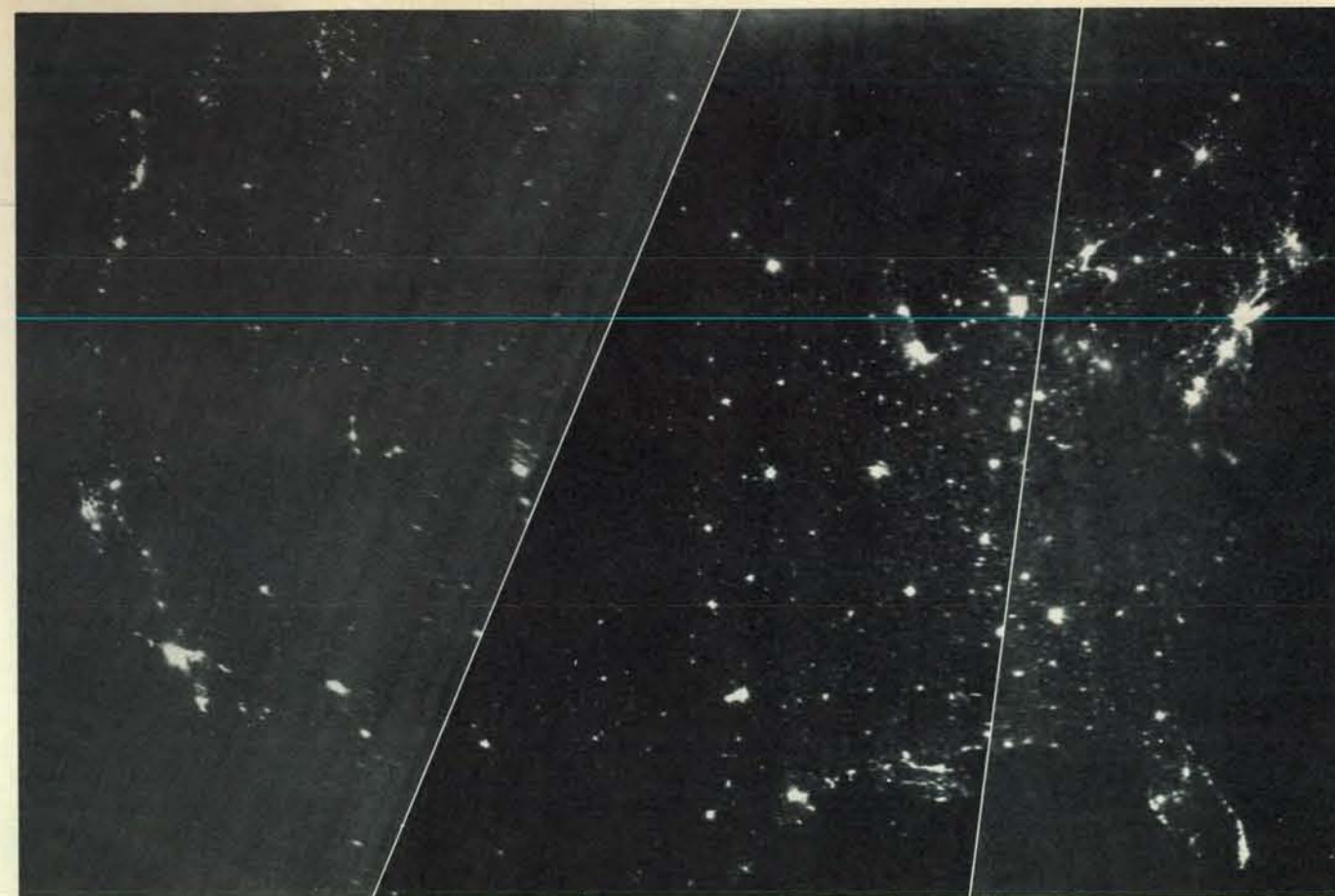



Immagine a quattro colori di due grandi fiammate di gas in Algeria, scattata di giorno dalla navicella spaziale *Landsat 1*, messa in funzione dalla National Aeronautics and Space Administration. Qui le fiamme del gas incendiato compaiono come macchie rosse in cima ai lunghi pennacchi di fumo. Alla pagina seguente in alto c'è una fotografia fortemente ingrandita della più meridionale di queste fiammate.

Ingrandimento di una parte piccolissima di foto notturna dal Landsat, che comprende una delle due fiammate di gas algerine illustrate nella fotografia alla pagina precedente in basso, fatto dall'autore nel tentativo di misurare l'estensione in larghezza della fiammata. Ogni immagine del Landsat è composta da milioni di piccoli elementi, detti *pixel*, che corrispondono sulla superficie terrestre ad aree rettangolari lunghe 79 metri e larghe 57. Da ogni pixel il sistema fotografico del Landsat ricava quattro distinte letture del sensore, una per ciascuna di quattro bande spettrali. In questo caso nell'immagine ingrandita c'erano solo 35 pixel illuminati e quindi è stato possibile riprodurre singolarmente ognuno dei rettangolini colorati per mezzo di normali tecniche fotografiche. È stato usato un metodo di riproduzione a colori falsati, per spostare nella parte visibile dello spettro le lunghezze d'onda, usate dal Landsat, dell'infrarosso e dell'infrarosso vicino. L'immagine che ne risulta indica che l'estensione della fiammata era di circa 500 metri, anche se potrebbe trattarsi di due fiammate distanti 200 metri l'una dall'altra.

Paragone fra le bande spettrali usate dal sistema a scansione multispettrale del Landsat e quelle usate da due versioni del sensore a bordo del satellite meteorologico dell'Air Force. Le sei curve vengono tutte rappresentate con lo stesso valore di picco, sebbene in effetti sia il vecchio sensore dell'Air Force (rappresentato dalla curva bianca tratteggiata a destra) che quello nuovo (rappresentato dalla curva bianca tratteggiata a sinistra) siano più sensibili di tutti i sensori del Landsat (rappresentati dalle quattro curve bianche continue). La fascia di colori che attraversa verticalmente il grafico nell'intervallo di lunghezze d'onda che va da circa 0,4 a 0,7 micrometri corrisponde alla parte visibile dello spettro elettromagnetico. Il confronto delle curve dei sensori di questi satelliti con lo spettro visibile spiega perché ambedue i satelliti siano adattissimi a rivelare gli incendi: gran parte dell'energia trasmessa da un incendio è a lunghezze d'onda alle quali l'occhio umano è insensibile.

La distribuzione della popolazione degli Stati Uniti viene rispecchiata dalla distribuzione delle luci urbane, illustrata dal montaggio di tre foto notturne da satellite, in alto. In basso è riprodotta una carta della popolazione, elaborata al calcolatore, fornita dal Bureau of the Census.

Una brillante aurora boreale, che si estende attraverso il Canada e si spinge a sud fino ai Grandi Laghi, è l'aspetto più evidente di questa fotografia, ripresa dal satellite meteorologico dell'Air Force la notte del 18 aprile 1974. La parte centrale è occupata dalle luci urbane degli Stati Uniti. Le luci più piccole sul lato occidentale del Messico, in basso a sinistra, sono fuochi agricoli; la fiammata di gas grande e brillante, in basso a destra, è associata al centro petrolifero di Reforma.

agricoli che corre attraverso l'Africa centrale a sud del Sahara. Un altro buon esempio è illustrato nell'immagine da satellite qui a sinistra. Siccome le luci sparse in tutta la zona occidentale del Messico sono assenti in altri periodi dell'anno, io suppongo che anch'esse debbano essere causate da incendi agricoli; le fiammate di gas di questa immagine, invece, nel corso dell'anno non mostrano grandi variazioni. Le intense fiammate di gas che si vedono in questa striscia sono associate ai giacimenti petroliferi di Reforma, sulla costa orientale del Messico. (Questa particolare striscia illustra anche magnificamente che l'aurora boreale può talvolta apparire anche molto a sud.)

Incendi agricoli simili si possono vedere in molte altre parti del mondo e in genere si può stabilire che sono stati appiccati dagli agricoltori o dagli allevatori di bestiame in base al fatto che in certe stagioni essi scompaiono. Le immagini diurne del Landsat sono utili per lo studio della vasta zona di incendi agricoli dell'Africa centrale perché le aree bruciate, nei passaggi successivi, appaiono annerite. In questo modo, con le immagini del Landsat, è stata misurata la superficie totale di terreno ripulito in questa maniera.

Si è visto che le uniche zone dove si hanno molti incendi agricoli sono in paesi poco sviluppati; presumibilmente tale pratica è legata alla mancanza di macchinari agricoli a motore per preparare il terreno alla semina e di conseguenza al loro posto, per togliere i residui del raccolto della stagione precedente, si usa il fuoco. Naturalmente certi prodotti agricoli richiedono il fuoco anche in circostanze in cui esistono altre alternative; ciononostante è abbastanza sorprendente vedere in che misura il fuoco è ancora in uso nei paesi che risentono della mancanza di moderne macchine agricole. Un secolo fa incendi simili vennero appiccati per contribuire a eliminare le vaste foreste che allora coprivano gli Stati Uniti centrali.

Incendi naturali si vedono anche nelle immagini da satellite dell'Air Force, per esempio nell'Australia nordoccidentale *(si veda l'illustrazione a pagina 56)*. A differenza degli incendi nelle terre coltivate, questi sono molto grandi. Due aspetti indicano che questi incendi sono accidentali: sono in regioni isolate, di solito non coltivate, e sono così grandi che l'uomo potrebbe controllarli solo con difficoltà, sempre ammesso che ci riesca. (D'altra parte è stato riportato che un tempo gli aborigeni bruciassero la boscaglia per conservare i pascoli per i canguri; forse questa pratica è ancora in uso su grande scala.) È chiaro che questi incendi uccidono gli animali selvatici; ho ricevuto una richiesta d'informazioni da parte dell'Australian Wildlife Service, che esprimeva interesse nell'uso di queste immagini da satellite per controllare gli incendi ricorrenti in quella regione. È chiaro che tale progetto è sia fattibile sia, dato l'accordo della nazione in questione, desiderabile per controllare i danni causati dai grandi incendi naturali. Fatta eccezione

per i momenti di intensa nuvolosità, gli incendi potrebbero venire registrati giornalmente su mappa.

Un'altra civiltà, che guardasse la Terra dallo spazio, non vi vedrebbe alcuna prova di carenza di energia; in effetti l'osservazione notturna mostrerebbe grandi sprechi, che dovrebbero indicare una sovrabbondanza di energia. Lo spreco deriva in parte dalla volontà dei paesi sviluppati (in particolare di quelli che di notte sono bene illuminati) di non dipendere dalle importazioni di energia, perché tale dipendenza viene percepita come una minaccia alla loro sicurezza nazionale e al loro benessere economico.

Le luci urbane del mondo sviluppato e le fiammate di gas del mondo meno sviluppato sono eventi relativamente recenti nella storia; appena cent'anni fa immagini notturne della Terra vista dallo spazio non avrebbero mostrato città né fiammate, e incendi agricoli molto più rari. Cosa mostrerà il lato buio della Terra fra altri cent'anni? Vorrei azzardare l'ipotesi che le zone abitabili saranno assai simili agli Stati Uniti di oggi, ma senza fiammate di gas. Il petrolio e il gas naturale saranno altamente apprezzati come prodotti chimici e verranno usati come combustibili solo in certi sistemi a prezzi da capogiro.

Perché gli operatori dei giacimenti petroliferi non arrestano lo spreco del gas incendiato oggi, installando delle apparecchiature per liquefarlo? La risposta a questa domanda, apparentemente ovvia, ha tre aspetti. Primo, effettivamente miliardi di dollari vengono investiti attualmente in apparecchiature per la liquefazione del gas e in navi per trasportare il gas liquefatto; la velocità e la scala di questi sforzi sono state ostacolate dai dirigenti dei paesi sviluppati, che hanno affermato la loro avversione all'acquisto di tale gas liquefatto per il timore di diventarne troppo dipendenti. Qui di nuovo il problema della bilancia dei pagamenti e la preoccupazione per la sicurezza nazionale hanno come risultato indiretto l'incendio di altro gas di scarico.

Anche un secondo fattore svolge un ruolo importante: gli attuali macchinari per la liquefazione del gas sono intrinsecamente economici solo quando l'impianto è grande. I piccoli centri di scarico vanno collegati a grandi centri, altrimenti probabilmente i loro gas non possono venire conservati in maniera economicamente conveniente. È necessario che vengano progettati impianti per la liquefazione del gas che siano piccoli, ma economici. Forse le immagini da satellite riveleranno che le fiammate piccole sono abbastanza numerose da garantire un coraggioso investimento per lo sviluppo di tali impianti.

Terzo fattore che ostacola la strada dell'uso produttivo dei gas di scarico è semplicemente il fatto che spesso molte fiammate servono per distruggere gas di scarico altamente corrosivi, per i quali non è ancora stato trovato alcun uso economicamente conveniente.

Per quanto riguarda gli incendi agricoli sembra probabile che questo tipo di fuoco a scala così grande asporti dal suolo sostanze importanti gettandole al vento. Per quanto si sappia che il fuoco è necessario alla sopravvivenza di certi sistemi ecologici, sembra una coincidenza eccessiva il fatto che le regioni a produzione agricola più bassa siano anche le regioni con più incendi agricoli. Forse l'uso del fuoco per la preparazione del terreno alla semina aveva più senso quando la popolazione non era così numerosa. Soluzioni alternative all'incendio, che conservino le sostanze nutritive presenti nel suolo, potrebbero avere chiaramente un'importanza mondiale, se si tiene presente l'estensione di questa pratica, come viene ampiamente documentata da queste immagini notturne riprese dai satelliti.

Le grandi macchie luminose di questa fotografia notturna da satellite dell'Australia occidentale sono probabilmente incendi naturali; infatti sono molto più grandi dei fuochi controllati che si vedono nelle zone coltivate. Le brevi strisce orizzontali luminose sulle nubi sono fulmini. In basso a sinistra, illuminato dalla Luna, si vede l'angolo sudoccidentale dell'Australia.

Cartagine e le sue monete

Da indagini archeologiche e numismatiche è emersa la funzione economica e politica che Cartagine svolse nelle province del suo impero, affiancando alla strategia bellica una articolata azione di penetrazione commerciale

di Enrico Acquaro

Lo scontro tra Roma e Cartagine nel Mediterraneo centrale sintetizza, nell'incalzare delle operazioni militari che coinvolgono popoli e regioni diverse, temi di opposizione di varia natura e origine. Accanto infatti alle piú note motivazioni etnico-culturali, care agli autori classici, la recente indagine archeologica e numismatica individua piú puntuali e storicizzati motivi politici, strategici ed economici. In quest'ambito, particolare valore di documentazione assume la politica monetale di Cartagine e il rinvenimento in Africa, in Italia e in Spagna di migliaia di monete riferibili a zecche cartaginesi. I rinvenimenti confermano, precisano e spesso individuano per la prima volta gli articolati ruoli economici e politici che Cartagine svolse nelle diverse province del suo impero, la sua competitività nei riguardi dell'economia prima siceliota e poi romana, la sua capacità di affiancare ad azioni belliche strategie di complessa programmazione politica, in cui ben predisposti sistemi alternativi di economia monetale, realizzati spesso nei gangli piú vitali della stessa confederazione italica, giustificano appieno, più del vago *metus Punicum,* i timori romani per la rivale africana.

La capacità perturbatrice del sistema italico, che Cartagine raggiunge in massimo grado all'epoca della discesa di Annibale, si fonda su quel complesso di rapporti che la città punica aveva intrattenuto in precedenza con le popolazioni etrusche, campane e magnogreche fin dall'inizio della sua politica, e che aveva mantenuto spesso in accordo con Roma. Se il sistema monetale come modulo di propaganda politico-economica e di alternativa di mercato costituisce per Cartagine un impegnativo tirocinio che inizia solo verso la fine del V secolo a.C. con le prime emissioni siciliane, è anche vero che tale sistema raggiunge in breve tempo un notevole respiro finanziario e una scaltrita capacità concorrenziale, al passo con il mutare delle circostanze e dei territori di impiego.

In Sicilia, dove la presenza di una monetazione altamente evoluta come quella siracusana e agrigentina richiedeva l'immissione in circolazione di monete che mantenessero, per quanto possibile, valori ponderali usuali e simboli noti o perlomeno suscettibili di immediate trasposizioni e assimilazioni, Cartagine affida in modo determinante la propria originale connotazione di emissione a leggende in punico o bilingui. Verso la fine del IV secolo a.C. la città africana riconduce l'esperienza monetale siciliana a parametri metropolitani di prestigio, in nulla inferiori alle più curate serie dei regni ellenistici. In Sardegna la mancanza di un mercato indigeno prima e la lotta con Roma poi determinarono emissioni che forse per la prima e ultima volta espressero a livello di economia evoluta la dignità nazionale del popolo sardo, anticipando di circa tre secoli la politica augustea volta alla valorizzazione nell'ambito dell'impero di miti, credenze e risorse tradizionali locali. Il rinvenimento di testimonianze numismatiche puniche nell'area etrusca sostanzia di significato commerciale ed economico i già noti rapporti culturali, politici e religiosi esistenti tra i punici e gli etruschi. La discesa di Annibale in Italia dietro l'esaltazione dei successi iberici imprime a tali rapporti una nuova e più diretta dinamica chiamando la Campania alla lotta contro Roma, in ragione anche degli antichi legami che i barcidi intrattenevano da tempo con la regione. E di questa politica la monetazione punica d'Iberia si fa portatrice nelle sue serie più prestigiose, largamente aperte alla propaganda istituzionale e politica della nuova aristocrazia di governo.

Questa prima esemplificazione della varietà documentaria che i rinvenimenti di monete puniche nel Mediterraneo sono in grado di offrire all'attenzione degli storici può trovare in ciascuna delle aree ricordate motivi d'interesse nella possibilità di valutare all'interno di ogni esperienza provinciale variazioni e fluttuazioni, di confermare dati delle fonti classiche ove sussistono, addirittura di costituire l'unica testimonianza di fenomeni politici ove manchino.

In Sicilia le serie puniche si pongono fin dall'inizio a livello competitivo rispetto alle coniazioni delle colonie greche. Nato a sostegno dell'azione colonizzatrice, l'impegno monetale di Cartagine, che ha corso regolare nel territorio metropolitano solo dal III secolo a.C., ha inizio nell'isola alla fine del V secolo a.C. Con la pace del 404 a.C., che delinea in Sicilia una situazione politica a vantaggio di Car-

Tetradramma d'argento di grammi 17,03/17,09 proveniente dalla zecca punica di Sicilia (320-300 a.C.). E' in Sicilia che Cartagine dà inizio alla propria attività monetale riprendendo tipi di diretta ispirazione siceliota. Al diritto della moneta *(a sinistra)*, appare la testa di Core circondata da delfini; al rovescio *(a destra)* è raffigurata la protome equina affiancata da palma e con l'espressione «popolo del campo».

Due stateri di elettro e oro rispettivamente di grammi 7,19 e 7,70 della zecca punica di Cartagine (350-320 a.C.). Verso la fine del IV secolo, Cartagine riconduce l'esperienza monetale siciliana a parametri metropolitani di prestigio, in nulla inferiori alle più curate serie dei regni ellenistici. Al dritto degli stateri è raffigurata la testa di Core *(in alto)*, al rovescio il cavallo stante a destra sulla linea di esergo.

Tetradramma d'argento di grammi 17,18 della zecca punica di Sicilia (320-300 a.C.). Le emissioni siciliane di Cartagine mantengono per tutta l'epoca della loro coniazione un eccellente livello di realizzazione.

Al dritto della moneta *(a sinistra)* si vede una testa femminile con berretto frigio; mentre al rovescio *(a destra)* compaiono il leone, con dietro la palma e, sotto l'esergo, l'espressione «popolo del campo».

tagine, la necessità d'inserirsi in un commercio dei più attivi ed evoluti dell'antichità, fondato sull'uso del metallo monetato, allinea le zecche dei centri punici di Mozia, Solunto ed Erice in un tipo di monetazione che ricalca da vicino le più prestigiose emissioni siceliote.

Al regime bimetallico, in bronzo e in argento, di Agrigento e di Selinunte si rifà in particolare la zecca di Mozia, attiva fra il V secolo a.C. e il 397 a.C., data della distruzione della città a opera di Dionisio di Siracusa. I tipi monetali registrano in alternanza e in commistione temi agrigentini, come l'aquila e il granchio, e i temi di antica tradizione punica, come la palma. La connessione di quest'ultimo tipo con il *gorgoneion*, in stretta analogia con alcune serie argentee di Populonia, individua un interessante motivo di contatto fra le due esperienze culturali ed economiche, la punica e l'etrusca. A questioni di praticità di mercato sembra si possa attribuire la coerenza metrologica della monetazione moziese con quella dei centri della Sicilia occidentale. La città rivela inoltre, in base agli ultimi rinvenimenti monetali, un rapporto commerciale diretto con Erice, Siracusa, Imera e Segesta. Le leggende in punico, che si alternano all'etnico greco, conoscono alcune varianti e diversa disposizione rispetto al tipo figurato.

La doppia leggenda in punico e in greco individua la monetazione punica di Solunto, che inizia anch'essa nel V secolo a.C. Fra i motivi figurati adottati sono in netta prevalenza quelli d'ispirazione greca: il gallo, il tonno, l'Eracle barbato e il toro androcefalo. Due sole emissioni documentano per Erice l'impiego della leggenda punica *'RK*. I tipi adottati sono la testa di Afrodite e il toro androcefalo, la testa di Atena con elmo corinzio e il pegaso alato. Questi ultimi motivi costituiscono una riprova diretta e significativa che l'influenza della monetazione corinzia, affermatasi nell'isola a seguito della spedizione di Timoleonte, non risparmiò neppure l'area punica.

Accanto alle tre zecche sicuramente operanti nell'ambito cittadino di Mozia, Solunto ed Erice, rimangono a testimonianza dell'inserimento attivo di Cartagine nella vita economica e culturale della Sicilia antica diverse serie, in argento e in bronzo, che si contrappongono al tentativo di risanamento monetale che Dionisio di Siracusa tentò di operare nella regione orientale, a seguito della violenta crisi economica provocata dalla spedizione contro Atene. L'esigenza di far seguire immediatamente alla conquista territoriale l'affermazione di prerogative economiche e amministrative è documentata in particolare da una delle più antiche serie enee emesse da Cartagine nell'isola, quella con i tipi del gallo e con la leggenda *ZYZ*. Sembra infatti che tale emissione sia da ascriversi all'attività di una zecca di Imera, rimessa in attività dai conquistatori cartaginesi dopo la distruzione della città calcidese da loro operata nel 409 a.C.

Tetradrammi argentei con la stessa leggenda, e di buona fattura, ricalcano nell'impostazione generale e nello stile le

Rovescio di un grande bronzo di 14,50 grammi, proveniente dalla zecca punica di Sardegna (264-241 a.C.), in cui è raffigurato un cavallo stante a destra, retrospiciente. Nel campo a sinistra, vi è la lettera *kaf* e due tratti (numero dodici). Il tipo è ribattuto su protome equina volta a destra. L'emissione di questa serie, caratterizzata per lo stile approssimato e per la tecnica poco curata, si riferisce alla profonda crisi economica che dovette colpire la Sardegna negli anni precedenti o subito seguenti l'inizio della guerra contro Roma. La presenza nell'isola di numerosi contingenti di mercenari e la precaria condizione economica danno origine a questa moneta, dalla ribattitura imperfetta e dalla circolazione limitata, presente nei maggiori ripostigli sardi.

Frazione di statere d'oro di grammi 1,75 proveniente dalla zecca punica di Cartagine (350-320 a.C.). Nel corso del III secolo le emissioni metropolitane, pur mantenendo tipi di stile e tematica sicelioti, si differenziano alquanto dai primi modelli, soprattutto nel dritto. In questo statere, al dritto *(a sinistra)*, è raffigurata la testa di Core; al rovescio *(a destra)* un cavallo stante a destra sulla linea di esergo.

Le due facce di un grande bronzo di 15,70 grammi della zecca punica di Sardegna (264-241 a.C.). L'inizio della monetazione bronzea in Sardegna fu determinato dall'accresciuto volume del mercato monetale locale e dalla necessità di integrare la circolazione delle monete siculo-puniche. Al dritto *(a sinistra)*, è raffigurata la testa di Core; al rovescio *(a destra)*, la protome equina con, nel campo a destra, la lettera *'ayin*.

più pregevoli emissioni siceliote, in particolare siracusane. Oggetto di un'attenta e sapiente resa tipologica che presuppone l'impiego di maestranze greche, libere o servili, le serie adottano sia tipi di diretta ispirazione siceliota, come il cavallo, la protome equina incoronata da Nike, la palma, la testa di Core con copricapo frigio e leone con palma.

Alla fine del IV secolo a.C., sempre su modelli sicelioti, ma con graduale affermazione di tipi punici, si daterebbe secondo alcuni l'inizio della coniazione nell'isola di serie auree cartaginesi. Le leggende puniche che appaiano sulle monete in metallo prezioso finora ricordate risultano spesso di difficile interpretazione per quanto riguarda il loro significato amministrativo. Accanto infatti a formule che sono diretta espressione della presenza militare di Cartagine nei territori siciliani, come «il popolo del campo», e altre che sembrano richiamare l'esistenza di una qualche magistratura monetale, come i «controllori», ve ne sono altre con un generico riferimento geografico e territoriale, che non hanno trovato finora adeguata e convincente identificazione nella geografia storica della Sicilia antica.

La mancanza di rinvenimenti in territorio africano di questi tipi di emissioni conferma da un lato la volontà e l'impegno esclusivo di Cartagine nell'isola, a ricerca di un'integrazione economica e di un sistema di scambi in metallo monetato sostanzialmente estranei alla propria economia metropolitana; dall'altro lato ribadisce, anche dal punto di vista della circolazione, l'appartenenza delle emissioni al sistema siceliota, cui le monete siculo-puniche si adeguano sia per stile sia per valori ponderali.

Ampia circolazione in tutto il territorio cartaginese, dalla Sardegna alla Spagna, hanno al contrario tre fra le numerose serie bronzee coniate con ogni probabilità da zecche puniche in Sicilia. Le monete, che portano al dritto costantemente il tipo della Core, hanno al rovescio il cavallo stante davanti a palma, la protome equina e il cavallo al galoppo o impennato a destra. Il rinvenimento a Selinunte di una notevole quantità di monete appartenenti alla prima serie insieme a metallo prezioso monetato e grezzo fa ritenere probabile, come già per Imera, l'attività nella città di un'officina monetale durante il lungo periodo dell'occupazione punica.

Due medi bronzi, il primo *(in alto)* di 3,50 grammi; il secondo *(in basso)* di 10,70 grammi, provenienti entrambi dalla zecca punica di Sardegna (241-238 a.C.). La serie cui appartengono queste due monete, con al dritto la testa di Core e al rovescio tre spighe, emessa durante la fase finale della prima guerra punica, costituisce una significativa concessione fatta all'ambiente locale e un estremo tentativo di Cartagine di chiamare alla compartecipazione nella lotta contro Roma anche la Sardegna. Nei rovesci delle due monete vi sono raffigurate tre spighe; sulla mediana si vedono il globo e la falce con i corni volti in basso; a destra la lettera *'ayin*; a sinistra, nella prima moneta, la lettera *dalet;* nella seconda, la lettera *res.*

Le immagini riproducono il doppio ciclo d'argento di grammi 14,61, della zecca punica di Spagna (237-218 a.C.). L'alto rendimento delle miniere d'argento iberiche e l'impegno politico e istituzionale dei barcidi trovano immediato riflesso in pregevoli emissioni d'argento. Al dritto della moneta è raffigurata una testa barbata, con mazza a sinistra (Amilcare?); al rovescio un elefante stante a destra con un conduttore.

Un interessante, tardo fenomeno di riconiazione è documentato a Cartagine su monete in bronzo di Gerone II di Siracusa. L'operazione, successiva alla perdita delle ricche miniere d'argento iberiche, mostra ancora una volta da parte dello stato cartaginese il ricorso al mercato monetale siceliota in un momento particolarmente delicato della propria storia, a riprova indiretta, ma concreta, della politica di equilibrio che Siracusa svolse in questo periodo fra Roma e Cartagine.

L'attività monetale che Cartagine fu indotta a iniziare in Sicilia costituì un impegno significativo per la capacità politico-amministrativa dello stato africano. La prova resa fu, per quanto risulta dai dati monetali, pronta e positiva, improntata a un realismo dei più produttivi che mostrò quanto Cartagine, a contatto con una realtà diversa dalla propria tradizione commerciale ed economica, si sia saputa adeguare, assumendo in veste più attuale e agile il ruolo di potenza colonizzatrice. Tale prontezza è indizio, d'altra parte, di un'antica consuetudine di rapporti con il mercato evoluto dell'Occidente mediterraneo, magnogreco ed etrusco, consuetudine che permette appunto a Cartagine un inserimento adeguato e naturale che sarebbe risultato altrimenti macchinoso, lento e certamente meno duttile. All'esperienza siciliana Cartagine fece costante riferimento negli altri suoi impegni monetali in Africa, in Sardegna, nella penisola italiana e in Iberia.

Le monete metropolitane, stateri tagliati su un peso di 9,5 grammi, portano al dritto una testa femminile che riprende, pur con alcune varianti, il tipo della Core già adottato nelle emissioni di Sicilia. Il tipo abituale del rovescio è il cavallo stante, con dietro una palma nelle frazioni maggiori: le minori portano su un lato una palma e sull'altro una protome equina. Nel corso del III secolo a.C. il piede si riduce a 7,6 grammi e il metallo è utilizzato in leghe sempre più ricche d'argento, fino a determinare emissioni in elettro.

Durante la prima guerra punica è emessa a Cartagine una serie che, soprattutto per la resa della testa femminile, si differenzia alquanto dallo stile siceliota delle prime monete; alla stessa epoca le emissioni in elettro si intensificano e al tipo del cavallo a rovescio si aggiunge il motivo del disco solare fiancheggiato da serpenti urei. La sconfitta del 241 a.C. lascia a Cartagine la possibilità di coniare solo in argento e in bronzo. Dopo gli acquisti barcidi in Spagna, che conferiscono nuovo impulso all'economia metallica cartaginese e registrano l'adozione di nuovi tipi, e dopo la pesante flessione immediatamente successiva a Zama, la monetazione di Cartagine conosce un nuovo, anche se breve, periodo di floridezza. In quest'ambito rientrano la coniazione di monete d'argento e d'oro: lo stile, come già quello di epoca barcide, è ormai decisamente improntato a tematiche ellenistiche e ha perso ogni connotazione di autonomia.

La storia delle emissioni puniche in Sardegna conosce aspetti diversi, che riflettono il mutamento diacronico dell'impegno cartaginese nell'isola a seguito soprattutto del conflitto con Roma. Interessata dalla circolazione di numerale siculo-punico fin dai primi del IV secolo a.C. come mostrano alcuni rinvenimenti tombali, la Sardegna punica dovette iniziare le proprie coniazioni enee intorno alla fine dello stesso secolo, adottando tipi e metrologia delle tre serie anepigrafi siciliane con al dritto la testa di Core e al rovescio la protome equina e il cavallo con o senza palma. L'inizio di questa monetazione fu determinato con ogni verosimiglianza dall'accresciuto volume del mercato monetale locale e dalla necessità di integrare adeguatamente la circolazione alquanto irregolare delle monete siculo-puniche. La mancanza di leggende compiute (sono presenti solo lettere isolate) e di continuità o convergenza di qualificanti connotazioni tecniche e stilistiche nell'ambito delle varie serie non consente, allo stato attuale dei rinvenimenti, di definire l'esistenza o meno di determinate zecche cittadine per tutto l'arco della monetazione sardo-punica.

Un successivo aspetto dell'impegno monetale cartaginese in Sardegna è documentato dalla serie con al dritto la testa di Core e al rovescio tre spighe. La serie, emessa durante la fase della prima guerra punica, e che presenta al rovescio un tipo del tutto nuovo nell'ambito del tradizionale repertorio monetale punico, costituisce una significativa concessione fatta all'ambiente locale e un riconoscimento esplicito del ruolo ben definito e apprezzato svolto dalla Sardegna nell'ambito dei possedimenti punici. In sostanza, l'emissione delle monete con spighe segna l'estremo tentativo da parte di Cartagine di chiamare alla compartecipazione nella lotta contro Roma anche l'isola, alla cui nota capacità produttiva cerealicola la metropoli africana attinse in più di un'occasione.

Le ultime due serie della monetazione sardo-punica, che portano l'una la testa di Core al dritto e il toro al rovescio, l'altra

Sempre della zecca punica di Spagna è questo triplo siclo d'argento di 22,05 grammi. Con i barcidi la monetazione punica di Spagna assume temi di diretta propaganda politica, istituzionale e religiosa. Nella moneta è raffigurata una testa virile volta a sinistra (Annibale?).

una testa virile imberbe e il toro, documentano una delle fasi più decisive della storia della Sardegna antica. Comunemente datate intorno al 216 a.C., e poste in connessione con la lunga serie di ribellioni e di repressioni che si susseguono nell'isola caduta ormai sotto il controllo romano e culminano nel 215 con la sfortunata insurrezione guidata da Ampsicora, le emissioni sembrano rientrare nella più vasta politica di riscossa che la famiglia barcide s'impegnò a condurre contro Roma. Le monete, coniate sotto gli auspici della battaglia di Canne e della defezione di Capua, riflettono un vasto disegno mediterraneo, in cui anche la monetazione, come nelle monarchie ellenistiche d'Oriente, assume temi di diretta propaganda politica. Da qui l'adozione del toro, senz'altro grato ai sardi per antica consuetudine di culto, ma anche non estraneo allo stesso mondo italico-campano che Annibale con la sua impresa in Italia stava chiamando a raccolta in funzione antiromana. Da ultimo, la testa apollinea che appare al dritto dell'ultima serie rende evidente l'influsso ellenistico dei barcidi. L'affiancamento a queste emissioni enee di altre degli stessi tipi in oro e in argento, le uniche registrate con certezza in Sardegna, danno un respiro mediterraneo alla politica monetale sardo-punica, che sembra uscire con esse dal suo isolamento, sotto la spinta rinnovatrice che la stessa famiglia condusse nello stato cartaginese corroborandone l'azione ed esaltandone la capacità competitiva. Se infatti la serie delle tre spighe costituisce il tentativo di rinsaldare, pur in un rapporto rinnovato, il legame già esistente fra territorio coloniale e metropoli, le serie con il toro inseriscono questo rapporto in un ambito più vasto che non sfugge alla suggestione di nuove forze politiche e istituzionali.

Alla stessa politica, ma con più esplicito collegamento con l'*etnos* italico, si rifà con ogni probalità l'emissione in Campania di monete in elettro battute sotto controllo punico. L'emissione, che ha al dritto la testa femminile bifronte e al rovescio Zeus su una quadriga guidata dalla Vittoria, è da localizzarsi con ogni verosimiglianza in una zecca di Capua. Parallela ad altre in argento e in bronzo con i tipi della testa di Core e del cavallo o della protome equina, attribuibili tutte a zecche dell'area campana e della zona meridionale della penisola, la serie di Capua è, come queste ultime, da collegarsi agli avvenimenti della campagna di Annibale in Italia fra il 218 e il 202 a.C. Ancora una volta la documentazione numismatica consente di meglio valutare i momenti storici con l'apporto di una testimonianza diretta: i famosi ozii di Capua, con tanta retorica ricordati dagli autori classici, appaiono, in accordo con la più recente critica storica, come un momento dei più importanti e delicati della campagna di Annibale in Italia, in cui vengono elaborate in concomitanza con le azioni militari corrosive strategie politico-economiche nell'intento di spezzare il già incrinato fronte italico.

Le coniazioni annibaliche, tuttavia, con la loro credibilità economica e politica, non dovettero risultare del tutto nuove alle popolazioni dell'Italia meridionale, raggiunte dalla circolazione di monete puniche fin dall'epoca del primo impegno cartaginese in Sicilia.

Da ultimo, la monetazione di tradizione punica di Pantelleria, la romana Cossura, che costituì un ganglio vitale della potenza militare cartaginese a controllo del Canale di Sicilia, propone un aspetto ancora diverso della presenza monetale punica nell'Italia antica. Analoghi alla monetazione maltese, i pochi numerali noti mostrano il perdurare di temi punici anche nelle successive coniazioni a leggenda latina. Motivo iconografico costante è al dritto un busto isiaco coronato da una Vittoria e al rovescio una corona di alloro o di mirto che iscrive la leggenda. La leggenda in punico reca il nome dell'isola, 'YRNM, del tutto dissimile da quello tramandato dalle fonti classiche e dalla successiva leggenda latina. L'inizio delle coniazioni, che solitamente si pone proprio intorno al 217 a.C., anno dell'occupazione romana dell'isola, testimonia la vitalità di una tradizione culturale punica anche in assenza del supporto politico e amministrativo: una tradizione che, sfumata dalle più generiche connotazioni egittizzanti della figura isiaca, assume per Roma l'aspetto rassicurante di un'innocua cultura locale, in cui il ricordo della competitività economica e politica della monetazione punica sembra superato. La cultura punica, con i suoi motivi figurativi e religiosi entra anch'essa, per singolare che sembri e sia pur limitatamente a un'esperienza provinciale e marginale, nel patrimonio di una cultura latamente italica che Roma eredita nella sua vocazione unificatrice.

In Iberia, l'alto rendimento delle miniere (trentasei tonnellate d'argento da una miniera chiamata Bebelo, cento chili al giorno da quelle nei dintorni di Cartagine) trova immediato riflesso nelle pregevoli emissioni d'argento che si pongono generalmente tra il 238 e il 206 a.C. Due sono le serie che vi si riconoscono, la prima facente capo a zecche fisse, come Cadice e Ibiza, la seconda riferentesi a zecche occasionali, come Ampurias e Sagunto. L'ipotesi, verosimile, ma non dimostrata, dell'attribuzione a Cartagena delle monete non altrimenti riferibili a ben precise zecche cittadine, dà ragione dell'appellativo «nazionale» con cui si è soliti designare i coni d'oro, d'argento e di bronzo che hanno come unità argentea il ciclo di 7,5 grammi. Le iconografie più usuali sono al dritto la testa di Core e quella di Eracle cinta d'alloro, al rovescio il cavallo, con o senza palma, e l'elefante. I tipi monetali di Cadice e di Ibiza riflettono una maggiore autonomia, accogliendo al dritto e al rovescio iconografie che esaltano i temi religiosi ed economici propri delle due comunità cittadine, quali la testa di Eracle, il tonno e il dio Bes.

Apprendimento e memoria nelle api

Un'ape è in grado di imparare rapidamente e di ricordare a lungo il colore e l'odore di fiori che forniscono nettare o polline. Oggi è possibile spiegare la base neurale di tale comportamento programmato

di Randolf Menzel e Jochen Erber

Se osserviamo un'ape mentre vola di fiore in fiore, vedremo che sceglie sempre il medesimo tipo di fiore e non dedica alcuna attenzione ai fiori vicini, di colore, figura e odore diversi, che pure attraggono altre api. È chiaro che l'ape non potrebbe presentare questo modello di comportamento se non avesse due precise capacità: quella di distinguere fra tipi di fiori diversi e quella di imparare quali fiori offrano nettare o polline e meritino quindi una visita. L'ape ha, quindi, memoria e capacità di apprendimento.

Noi e alcuni nostri colleghi della Freie Universität di Berlino abbiamo studiato queste capacità dell'ape nella speranza di gettar luce sulla base cellulare dell'apprendimento e della memoria. I nostri studi riguardano tanto il comportamento quanto la fisiologia. La nostra conclusione è che, benché l'ape sia una macchina comportamentale programmata in modo molto preciso, i suoi meccanismi di apprendimento e di memoria sono molto simili a quelli operanti in organismi molto più evoluti, compreso l'uomo.

Benché fosse ovvio a molti attenti osservatori del passato (i commenti scritti che ci sono pervenuti vanno da Aristotele a Charles Darwin) che l'ape è in grado di vedere il colore dei fiori e sentirne l'odore e di ricordare il tipo di fiore visitato in precedenza, la dimostrazione sperimentale di queste sue capacità venne solo nel secondo decennio del secolo, quando Karl von Frisch, all'Università di Monaco di Baviera, eseguì un semplice esperimento che è oggi considerato uno fra i pochi veramente cruciali e determinanti nello studio del comportamento animale. Egli attrasse delle api bottinatrici a un tavolo offrendo loro una soluzione di miele. Quando un numero di api abbastanza grande ebbe preso l'abitudine di visitare regolarmente il posto, von Frisch collocò il piatto contenente il miele su un pezzo di cartone blu, in modo che le api vedessero il colore blu quando si avvicinavano al tavolo e quando succhiavano il miele.

Karl von Frisch era ora pronto per il suo esperimento. Egli sistemò tutto attorno al cartone blu altri pezzi di cartone della stessa grandezza, ma di colore variabile fra il bianco e il nero, passando per varie gradazioni di grigio. Su ciascun pezzo di cartone, compreso quello blu, pose un piatto vuoto. Le api, che erano state condizionate ad attendersi di trovare cibo sul tavolo, continuarono a cercare, e quasi tutte limitarono la loro ricerca al cartone blu.

L'esperimento fornì un forte sostegno a entrambi i postulati da noi menzionati. Le api distinguevano il colore blu dagli altri colori: ciò significa che erano in grado di vedere il blu come colore e anche che avevano appreso a considerare il colore (in tale situazione) come un segnale del cibo. Per confermare questa seconda conclusione era necessario dimostrare che le api non venivano attratte spontaneamente da quel colore. Von Frisch lo dimostrò, e la cosa fu confermata da vari altri ricercatori. Il suo esperimento è stato in effetti ripetuto con una varietà di colori, odori e disegni in bianco e nero.

Gli aspetti comportamentali dell'attività delle api sono perciò ben noti. Sorprende invece che il processo di apprendimento implicito in tutti questi esperimenti non sia stato studiato fino a una decina di anni or sono. Noi dimostreremo che tale processo ha vari caratteri interessanti, che lo fanno considerare importante ai fini di un approccio a un'analisi della formazione della memoria nel cervello dell'ape su base nervosa.

Poniamoci una domanda. Quale grado di precisione dimostra l'ape nello scegliere il tipo di fiore da cui raccogliere il nettare e il polline? A questa domanda si può rispondere rapidamente e con esattezza senza bisogno di seguire un'ape nel suo volo di fiore in fiore. Quando un'ape ritorna periodicamente all'alveare, trasporta un elenco delle sue scelte nei due mucchietti di polline rimasto aderente ai «cestelli» delle zampe posteriori. Esaminando il contenuto dei mucchietti, si vede che il polline portato nell'alveare da un'ape dopo un'uscita proviene per intero o quasi per intero da un solo tipo di fiore. Poiché un'ape visita in un'uscita fino a 500 fiori, il grado di precisione nelle scelte di un'ape è davvero molto alto.

Questa precisione di risposta, unitamente al fatto che l'ape può essere addestrata facilmente a rispondere a vari stimoli, rende questo insetto particolarmente idoneo agli esperimenti che avevamo in mente. Risulta molto utile anche il fatto che tutte le api operaie di un alveare siano sorelle e quindi strettamente legate dal punto di vista genetico, e altrettanto utile è la loro forte motivazione alla ricerca del cibo, cosicché ciascuna di esse prende un gran numero di decisioni (da 1500 a 2000 al giorno).

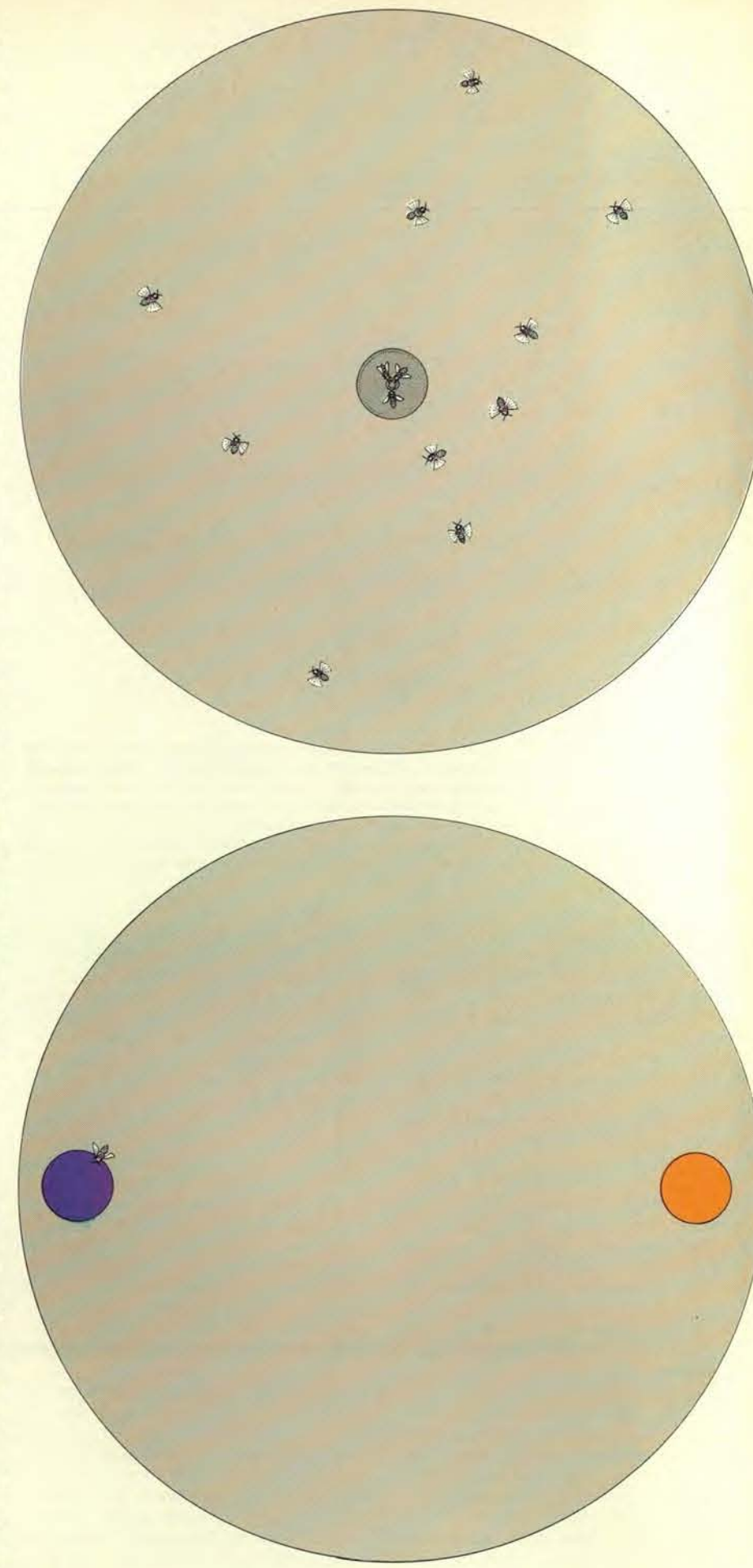
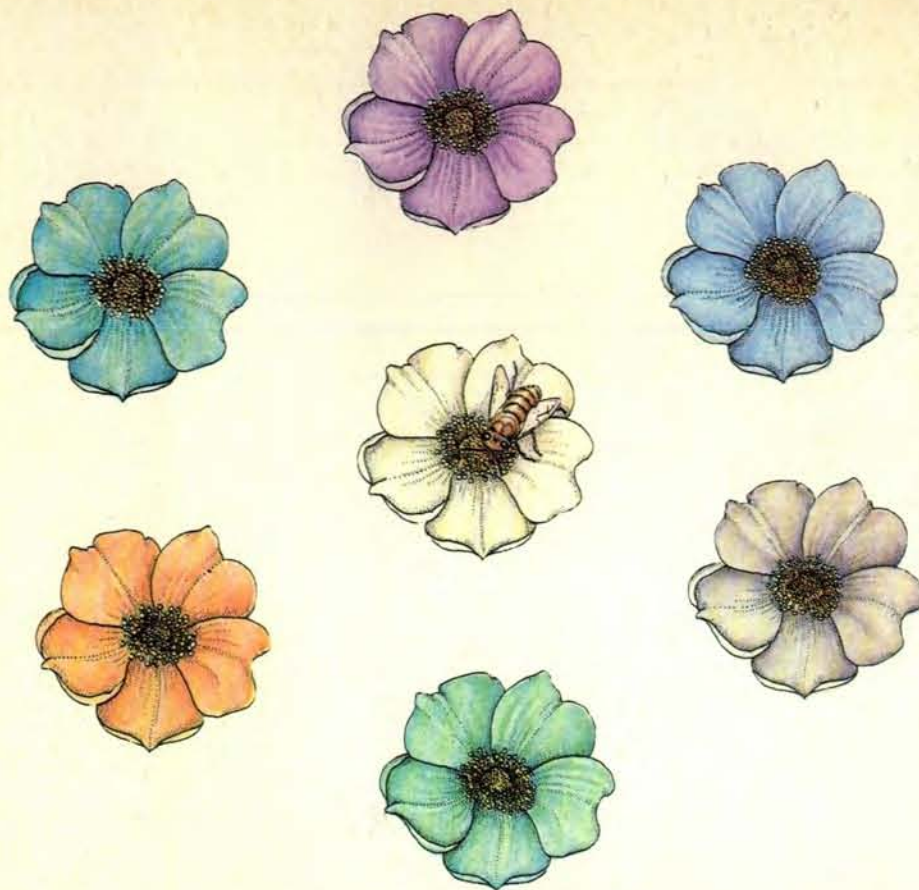
Per prepararci a un approccio diretto agli eventi nervosi sottostanti all'apprendimento e alla memoria, esaminammo dapprima dettagliatamente il processo di apprendimento con metodi comportamentali, concentrando particolarmente la nostra attenzione su esperimenti in cui le api venivano addestrate a rispondere a vari colori. (Altri ricercatori hanno adottato impostazioni diverse. Per esempio, Martin Lindauer e i suoi colleghi all'Università di Würzburg addestrarono le api o a odori particolari o a un programma orario giornaliero e Rüdiger Wehner, dell'Università di Zurigo, insegnò loro a distinguere fra vari disegni in bianco e nero.)

Le nostre apparecchiature per l'addestramento delle api al colore consistevano in un dispositivo che poteva usare luce monocromatica di intensità variabile per illuminare due dischi di vetro. Delle api marcate singolarmente ricevono una soluzione zuccherina su un disco non illuminato posto al centro del tavolo grigio usato per l'addestramento. Per mezzo di una tipica danza (la cui spiegazione è dovuta allo stesso von Frisch) le api informano le loro compagne di alveare di questa fonte di cibo. A una nuova venuta che arrivava in risposta all'informazione consentivamo di visitare tre volte un disco non illuminato per farle acquisire familiarità con l'apparecchiatura usata. Dopo tre visite nel corso delle quali ha ricevuto del cibo, un'ape è motivata alla ricerca anche quando non gliene viene offerto. Durante le prime tre visite di una nuova ape, tutte le altre venivano allontanate dall'apparecchiatura e tenute in una gabbia di rete, in modo che solo l'ape studiata potesse volare avanti e indietro fra l'alveare e l'apparecchiatura.

Il nostro primo test valutava le preferenze che emergevano nella scelta spontanea dell'ape. La prima volta l'ape studiata nell'esperimento veniva posta di fronte a due dischi di vetro di diverso colore. Nessuno dei due dischi conteneva cibo ed entrambi erano equidistanti dal centro del tavolo (dove era stato collocato in precedenza il disco col cibo). L'ape si avvicina ai dischi, si posa su di essi e cerca il cibo. Se la frequenza della scelta non è uguale per i due colori, l'intensità di un colore rispetto all'altro viene modificata finché ogni preferenza di scelta scompare.

Dopo un periodo di prova di alcuni minuti, l'ape viene gratificata con del cibo posato in uno dei due dischi di diverso colore. Quando l'ape ritorna, viene sottoposta a un test con le stesse modalità già illustrate nel caso delle prove di preferenza spontanea. L'ape viene poi nuovamente gratificata, e così via. Procedendo in questo modo siamo in grado di ottenere curve di apprendimento per singole api e possiamo perciò dare una valutazione quantitativa dei fattori che influiscono sul comportamento di apprendimento delle api.

I colori vengono appresi dopo poche prove. Il viola (alle lunghezze d'onda da 400 a 420 nanometri) è quello che viene appreso più rapidamente di tutti: una sola ricompensa è sufficiente a produrre un'alta frequenza di scelte corrette. Il colore che viene appreso più lentamente è invece il blu-verde (490 nanometri). Le differenze tipiche delle curve di apprendimento sono indipendenti da fattori come l'intensità dei due colori (se l'intensità è di almeno mezza unità logaritmica al di sopra della soglia di percezione delle api), la lunghezza d'onda del colore alternativo, il tipo di preaddestramento, l'età dell'ape sottoposta al test, l'ora del giorno o la stagione e le condizioni meteorologiche.

Gli autori di questo articolo hanno addestrato le api all'apprendimento di certi colori per mezzo di un dispositivo che consentiva di illuminare piccoli dischi di vetro (di sette centimetri di diametro) con luce monocromatica di intensità regolabile. Dapprima alcune api marcate individualmente ricevono del cibo *(in alto)* su un disco non illuminato al centro del tavolo grigio sul quale si svolge l'addestramento. Con una tipica danza esse informano altre api dell'alveare sulla posizione della fonte di cibo. Una singola ape riceve del cibo sul disco centrale, viene sottoposta a test per accertare la sua risposta nella scelta spontanea di due colori equidistanti dal disco centrale *(in basso)* e poi viene gratificata attraverso una somministrazione di cibo sul disco centrale, illuminato ora con uno dei due colori. Anche una gratificazione breve e di scarsa entità motiva l'ape a tornare con costanza a quel particolare colore.

L'apprendimento di un colore da parte delle api è abbastanza rapido per alcuni colori e un po' più lento per altri, benché un'ape possa imparare qualsiasi colore risulti visibile per essa. In questa serie di disegni il fiore della rosa canina è rappresentato al centro come lo vede l'occhio umano e poi, in senso orario dall'alto, secondo la rapidità con cui l'ape è in grado di associare un colore a una gratificazione di cibo. Il colore che viene appreso più rapidamente è un violetto molto saturo, poi il blu, l'ultravioletto *(qui presentato come grigio)*, il verde e l'arancione. Il blu-verde, che per le api è il normale colore di fondo, è quello che, fra tutti i colori, viene appreso più lentamente.

Ci si attenderebbe una dipendenza dell'apprendimento dalla lunghezza d'onda, a dimostrazione di una stretta correlazione col sistema di visione cromatica delle api. Ma non è così. Il sistema di visione del colore nelle api è tricromatico, con recettori sensibilissimi alla radiazione ultravioletta (350 nanometri), alla luce azzurra (440 nanometri) e alla luce verde (540 nanometri). Per le api l'ultravioletto è il colore più brillante e più saturo, come dimostrò vent'anni or sono Karl Daumer, dell'Università di Monaco di Baviera. Come ci si può aspettare da un tale sistema, l'ape distingue con maggiore facilità il violetto e il blu-verde, poiché la sensibilità dei suoi recettori è particolarmente pronunciata in tale regioni; qui c'è inoltre una sovrapposizione nella sensibilità dei vari recettori. In effetti quest'ipotesi fu dimostrata corretta da Otto von Helversen, dell'Università di Freiburg (in Brisgovia). Tali sensibilità differenziali sono necessarie per una precisa identificazione dei fiori nei quali predominano i colori viola e blu, i quali sono specializzati nell'attrarre gli insetti per l'impollinazione incrociata entomofila.

Benché le api siano molto sensibili alla radiazione ultravioletta, non è questo il colore che apprendono meglio. Per le api la radiazione ultravioletta è la radiazione del cielo e pare che esse non si attendano di trovarla come colore puro nelle fonti di cibo. L'ultravioletto viene appreso molto più rapidamente se è mescolato con luce di grande lunghezza d'onda a formare il «violetto delle api». Il blu-verde è il colore di fondo dell'ambiente in cui operano le api e non ci si attende di trovarlo come segnale di cibo. L'evoluzione interattiva dei fiori e dei loro insetti impollinatori ha condotto non solo a un adattamento fra colori dei fiori e percezione del colore delle api, ma anche a predisposizioni di apprendimento nelle api. Questa interpretazione trova conferma in studi di Lindauer sull'apprendimento dell'odore nelle api; i colori che si presentano in natura vengono appresi più rapidamente di colori artificiali e insoliti.

Osserviamo, in questo contesto, la tendenza recente fra gli psicologi sperimentali ad apprezzare il valore del punto di vista etologico, secondo il quale il comportamento viene studiato nel modo migliore attraverso un riferimento costante all'ambiente e all'evoluzione naturali. L'apprendimento del colore e quello dell'odore nelle api sono fra le forme di comportamento che possono essere esaminate nel modo più fecondo in questo modo. Essendo un etologo, von Frisch interpretò intuitivamente i suoi risultati in questa prospettiva già 40 anni or sono. Oggi possiamo separare quantitativamente lo schema del sistema percettivo dell'ape (il programma comportamentale per la raccolta di informazione specifica al contesto) dai programmi nervosi per l'elaborazione di questa informazione e per la sua introduzione nella memoria.

Quale stabilità ha il ricordo di segnali di cibo nell'ape? Relazioni aneddotiche fornite da apicoltori suggeriscono che sia molto stabile e anche duraturo. Una fra le prime osservazioni quantitative fu eseguita nel 1963 da Lindauer, il quale trovò che le api arrivavano spontaneamente in un posto in cui era stato loro fornito cibo dopo cinque mesi di assenza durante l'inverno. Per ottenere una valutazione ancor più esatta quantitativamente, addestrammo delle api al blu o al giallo con un numero di gratificazioni da una a tre, le tenemmo in gabbie di rete singole nell'alveare per periodi di tempo variabili e poi sottoponemmo a test il loro comportamento di scelta. Dopo una gratificazione, la frequenza di una determinata scelta rimaneva alta per varie ore ma declinava il giorno seguente e raggiungeva il livello della scelta spontanea in un numero di giorni compreso fra uno e cinque. Dopo tre gratificazioni non riscontrammo però alcuna flessione nel numero delle scelte corrette per il periodo di circa 13 giorni in cui un'ape può sopravvivere in una gabbia. È evidente che poche gratificazioni sono sufficienti per fissare nelle api un ricordo molto stabile.

Diversa è la situazione se alle api viene consentito un nuovo apprendimento. Esse passeranno allora a una nuova fonte di cibo marcata con un colore diverso, ma la rapidità del passaggio dipende dal numero di gratificazioni che ricevettero nell'apprendimento del colore ora non più premiato. In questa situazione si può osservare un curioso comportamento ben noto da studi sui vertebrati. L'animale modifica più rapidamente il suo comportamento se ha avuto un lungo periodo di addestramento (25 gratificazioni o più) al colore iniziale. Un'altra scoperta è che, se due segnali sono associati a gratificazioni con frequenze diverse, l'animale sceglie secondo le probabilità di gratificazione. (Quest'effetto è stato studiato in modo particolareggiato da Bernhard Schnetter dell'Università di Würzburg.) Un altro esempio è l'addestramento simultaneo a più di un colore. L'ape può imparare contemporaneamente due colori e può distinguerli da altri colori. Se però vengono premiati simultaneamente più di due colori, l'ape smette di distinguere i colori. Tutto ciò rivela qualcosa sulla struttura logica che è alla base dei suoi programmi di apprendimento.

Tali effetti complessi e duraturi dell'apprendimento sono interessanti dai punti di vista dell'etologia e della psicologia comparata, ma non si prestano a un'a-

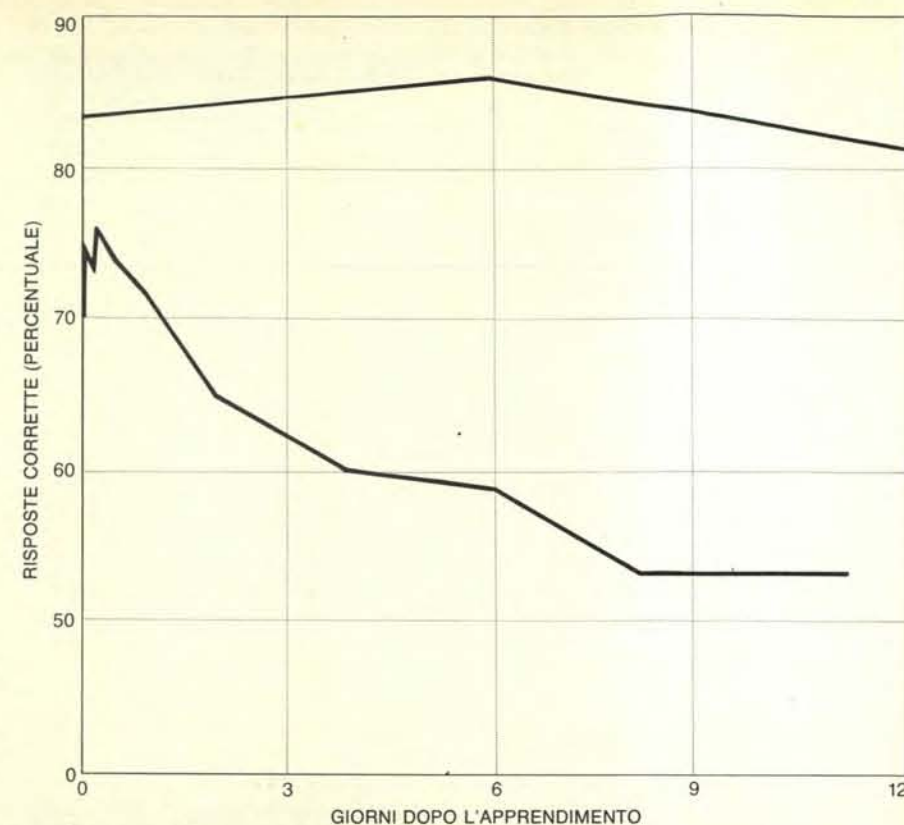
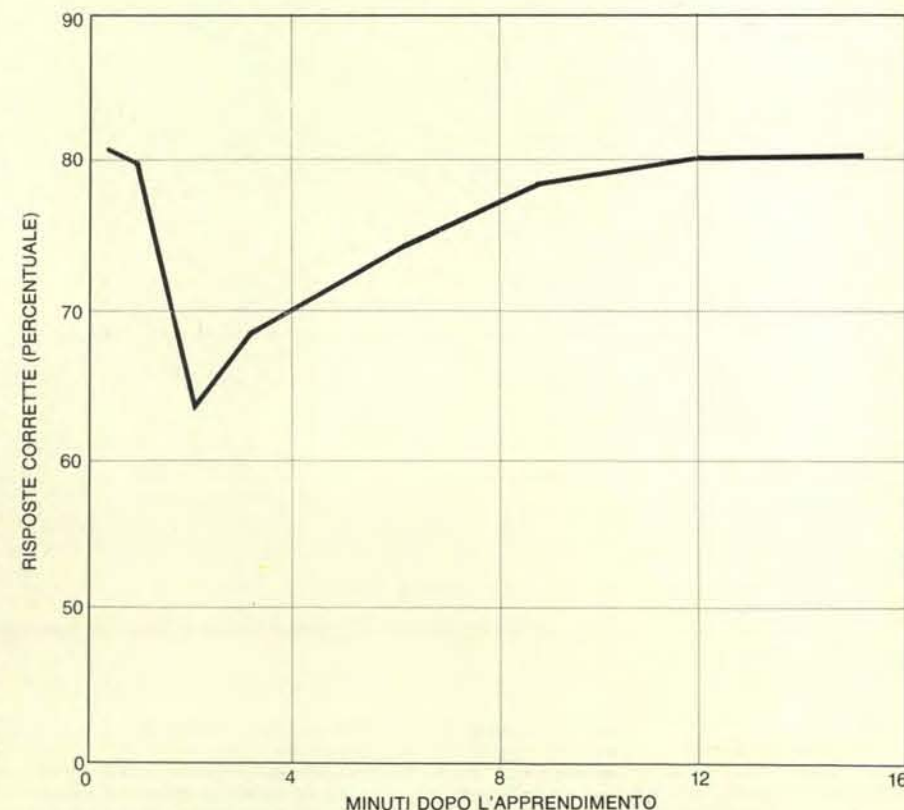
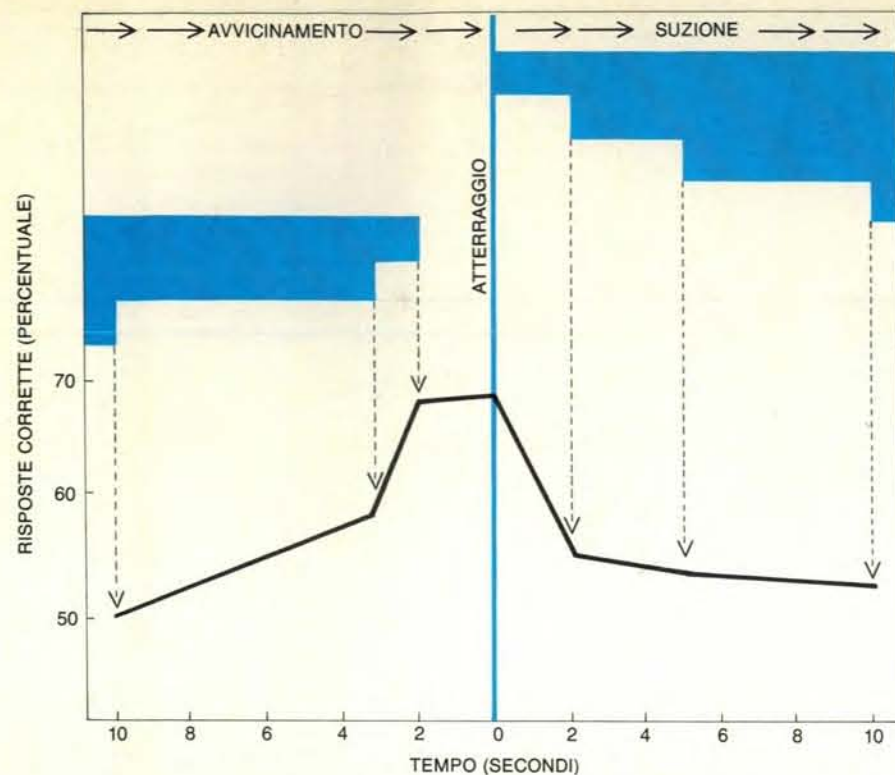
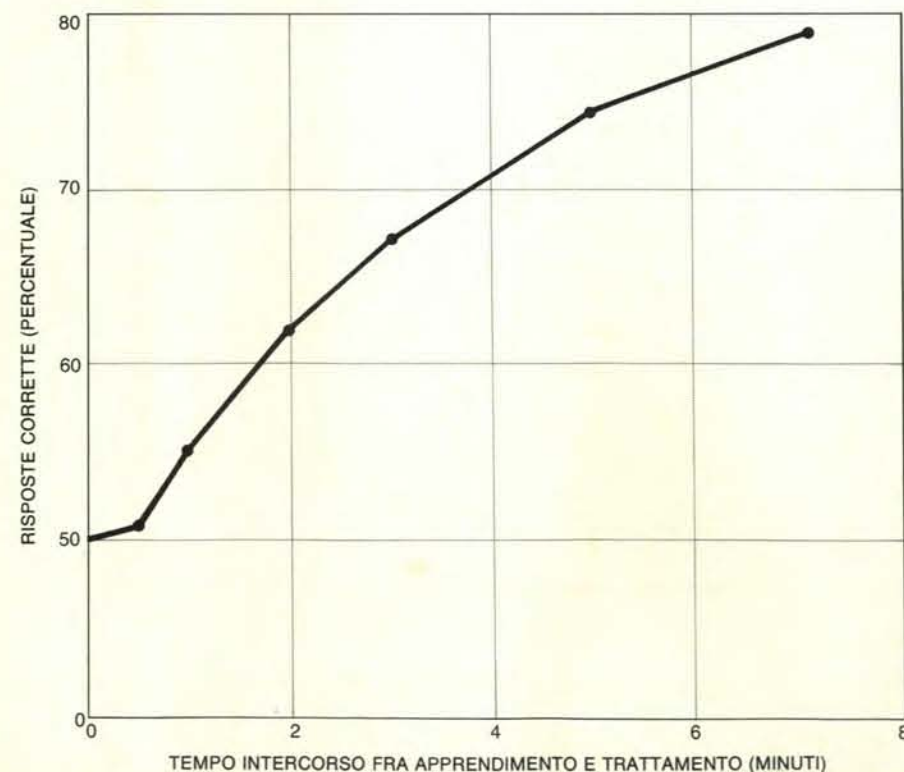
dimostra un livello di apprendimento significativo già dopo un singolo periodo di succhiamento di due secondi, il periodo totale di apprendimento iniziale (memoria sensoriale e periodo di associazione più la durata della gratificazione) risulta essere da quattro a cinque secondi. Questa è una situazione favorevole all'investigazione delle fasi temporali di un immagazzinamento del ricordo iniziato da una singola gratificazione.

Le fasi a breve termine e a lungo termine riflettono due meccanismi di fissazione fisiologicamente diversi? Abbiamo trattato le api con vari procedimenti noti per bloccare l'attività dei neuroni nel cervello. Tutti questi procedimenti inibiscono la formazione della memoria a breve termine. Le api trattate in questo modo si comportano, una volta tornate allo stato normale, esattamente come le api non trattate. Prove di controllo esclusero inoltre la possibilità che i procedimenti sperimentali causassero un rinforzo negativo.

Abbiamo trovato una correlazione fra il periodo iniziale di declino della memoria è il periodo di suscettibilità ai procedimenti bloccanti. I risultati ci dicono che, durante un periodo iniziale di alcuni minuti, la fissazione del ricordo nella memoria si fonda su meccanismi fisiologici diversi da quelli che hanno parte nella fissazione posteriore. È molto probabile che i primi meccanismi fisiologici siano attività nervose ordinate, dal momento che il nostro trattamento blocca tali attività.

Il primo passo nell'apprendimento associativo è chiaramente un meccanismo di immagazzinamento provvisorio dell'informazione sensoriale, fase che abbiamo chiamato di immagazzinamento sensoriale. Se l'informazione non è rinforzata prontamente, scompare. Il rinforzo del segnale visivo dipende da un secondo sistema sensoriale: i recettori dello zucchero sulla lingua dell'ape. Se i recettori sono stimolati da cibo, viene messa in moto la fase successiva del programma di apprendimento (immagazzinamento dell'informazione in un sistema di memoria).

È plausibile supporre che questi processi iniziali e rapidi di incanalamento e di selezione dell'informazione si fondino sull'attività elettrica di particolari cellule nervose. Registrazioni elettrofisiologiche potrebbero esserci d'aiuto nell'identificazione delle cellule interessate. Ma dove, nel cervello di un'ape, si dovrebbe inserire un elettrodo per la registrazione? Benché il cervello dell'ape sia relativamente piccolo (con un volume di un millimetro cubo circa), contiene qualcosa come 850 000 cellule nervose: troppe per una strategia di ricerca casuale. Perciò abbiamo preferito cercare di localizzare le aree del cervello implicate nei processi di fissazione iniziale del ricordo.

Il nostro metodo si fonda sulla risposta riflessa della proboscide a stimoli forniti da una soluzione zuccherina. Un'ape estende la proboscide non appena una sua antenna viene toccata da una soluzione zuccherina. Fissiamo un'ape in un tubo di metallo e le presentiamo un soffio di odore subito prima di toccarle un'antenna

In questo diagramma è rappresentata la stabilità della memoria delle api. Se le api ricevono tre gratificazioni in associazione con un particolare colore e poi vengono tenute nell'alveare fino a due settimane per escludere la possibilità di un nuovo apprendimento, mantengono in seguito un alto livello di scelte corrette *(curva in alto)*. Nel caso di una sola gratificazione, il livello delle risposte positive dapprima sale, nel corso di un periodo noto come fase di consolidamento. Il giorno dopo, però, la risposta comincia a calare e dopo circa sei giorni scende al livello della scelta spontanea.

Il comportamento di scelta delle api dopo un periodo di apprendimento in una sola prova presentò il seguente andamento: subito dopo l'apprendimento le api scelsero con precisione il colore premiato. Due minuti dopo il numero delle risposte corrette diminuì per tornare però a salire.

Il periodo di associazione fra un colore e la gratificazione è ben definito e breve. Un colore, che in questi esperimenti era il blu, visto dall'ape più di quattro secondi prima di posarsi sul cibo o più di due secondi dopo essersi posata e avere cominciato a succhiare, non viene associato alla gratificazione. Un segnale cromatico tolto due secondi prima dell'inizio della gratificazione viene invece appreso e, quindi, dovrebbe esserci una memoria sensoriale che persiste per almeno tre secondi.

Un procedimento di disturbo produsse gli effetti qui registrati sulle primissime fasi della formazione della memoria nelle api. Una volta determinata la scelta cromatica spontanea di un'ape, l'insetto fu gratificato in presenza di luce blu per 30 secondi. Poi il suo cervello fu raffreddato o immediatamente o dopo vari intervalli, indicati dai punti. La curva dimostra la percentuale media delle risposte corrette dopo il ritorno dell'ape sottoposta a trattamento alla normalità. Le prestazioni delle api sottoposte a raffreddamento furono dopo sette minuti press'a poco normali.

con la soluzione zuccherina e di consentire all'ape di lambire per alcuni secondi la soluzione della proboscide. Se l'odore viene poi presentato in assenza di una soluzione zuccherina, l'ape estenderà quasi sempre la proboscide. Una sola prova di condizionamento è sufficiente per stabilire un alto livello di risposte.

Trovammo che potevamo esporre il cervello durante questo processo senza influire in modo eccessivo sul comportamento di apprendimento. Potevamo poi bloccare in modo reversibile l'attività dei neuroni in piccole aree del cervello per mezzo di un sottile ago raffreddato. L'area bloccata ha un diametro di 250 micrometri e una profondità di 150.

Attraverso questo procedimento il nostro interesse venne a concentrarsi sui *corpora peduncolata*, noti anche come corpi fungiformi. Ogni ape ne ha due, fittamente intessuti di neuropili: il calice, a forma di bicchiere, l'antenna, il lobo frontale alfa e il lobo inferiore beta. Le informazioni principali gli provengono da un'altra struttura, il lobo antennale. Gli istologi dell'Ottocento osservarono che i corpi peduncolati sono particolarmente prominenti negli insetti sociali e ne inferirono che le funzioni superiori del cervello (apprendimento, memoria e comportamento sociale complesso) sono probabilmente localizzate in essi.

Quando menomammo parti dei lobi antennali e dei corpi peduncolati raffreddandole, trovammo una correlazione con quanto avevamo osservato bloccando l'attività nervosa in api lasciate volare liberamente. L'effetto della menomazione diminuisce con l'aumentare dell'intervallo di tempo compreso fra la prova e il trattamento. La suscettibilità alla menomazione diminuisce con la rapidità massima nel lobo antennale e con la minima nel calice.

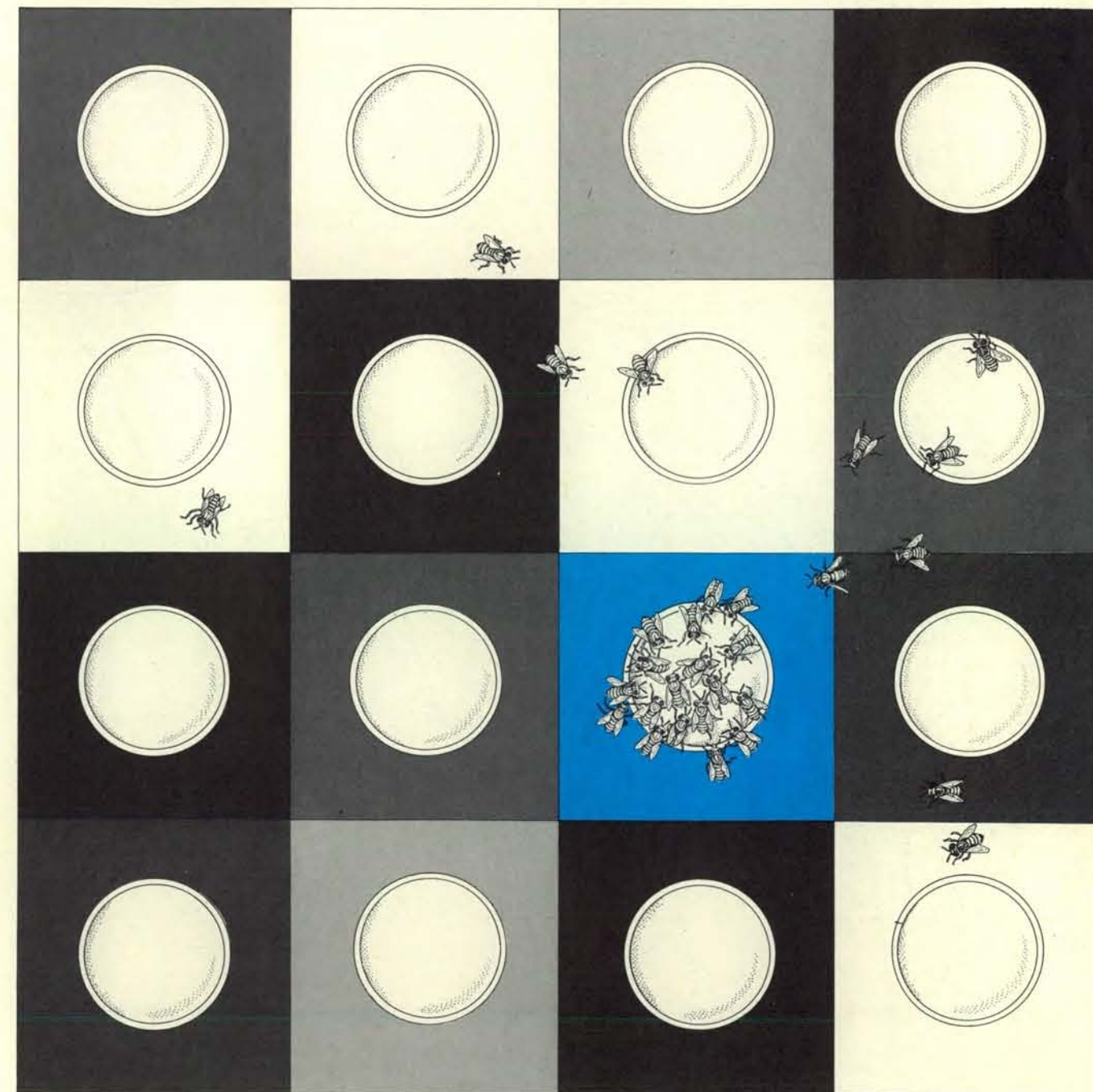
I risultati indicano che i lobi antennali e parti dei corpi peduncolati partecipano all'elaborazione dell'informazione nella memoria, anche se in misura diversa e per periodi di tempo diversi. I lobi antennali vengono coinvolti nel processo solo per un breve periodo iniziale di tre minuti circa. Il calice viene coinvolto più a lungo. Poiché si può dimostrare che la memoria viene menomata già sette minuti dopo il raffreddamento dei calici, questi sono probabilmente le strutture essenziali nel trasferimento dell'informazione dalla memoria a breve termine a quella a lungo termine. In questo processo hanno parte anche i lobi alfa, che sono connessi ai calici da migliaia di fibre nervose.

La situazione di apprendimento fondata su una sola prova adottata nei nostri esperimenti ci consentì di seguire il decorso temporale della trascrizione dalla memoria a breve termine in quella a lungo termine. I decorsi temporali che trovammo per varie strutture cerebrali ci fecero considerare estremamente improbabile che i processi elettrochimici e biochimici che partecipano alla trascrizione seguano una dinamica semplice all'interno di un'area sinaptica limitata. Dal

nalisi degli eventi nervosi sottostanti al comportamento. Durante e dopo una singola associazione di un segnale cromatico o olfattivo e una gratificazione concentrammo perciò la nostra attenzione sull'analisi dei processi nervosi. Questa si rivelò una decisione fortunata, perché improvvisamente emersero nuove prospettive in cui considerare il processo di apprendimento.

Da principio stabilimmo un risultato molto utile: dopo una singola gratificazione, il livello di risposta è indipendente dall'entità quantitativa della gratificazione, purché vengano osservate certe limitazioni di soglia. Persino una gratificazione che persista per un solo decimo di secondo e che abbia una bassa concentrazione di saccarosio è sufficiente a produrre un mutamento significativo nella risposta.

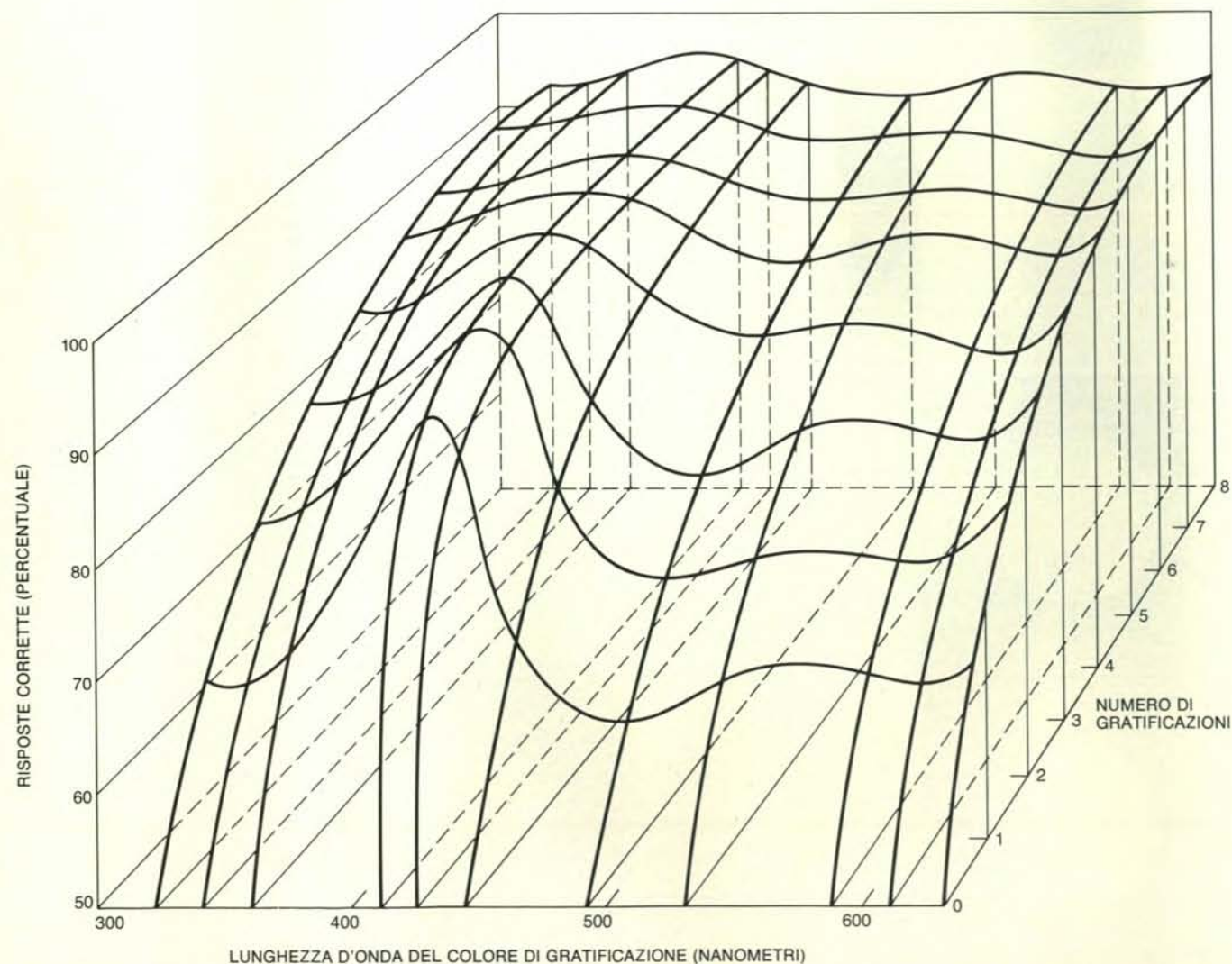
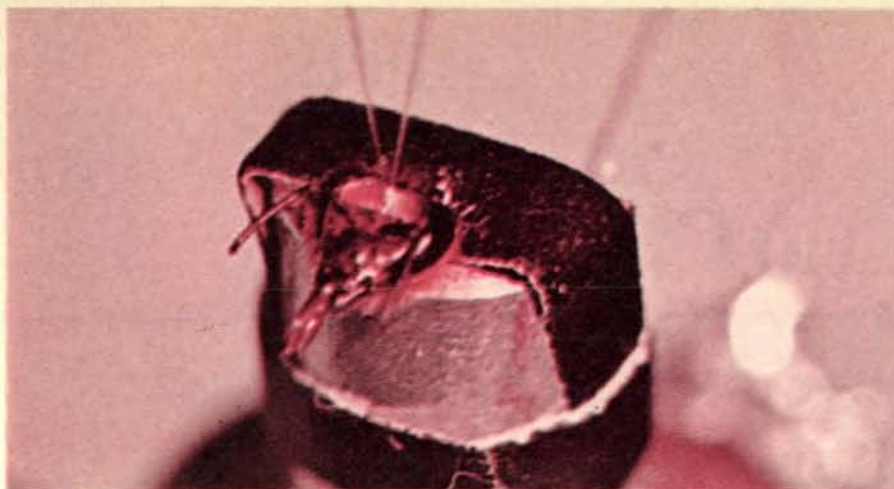
Un altro risultato interessante emerse quando addestrammo le api con una sola ricompensa breve (due, cinque o dieci secondi). Le api sottoposte a test subito dopo la ricompensa presentarono un'alta frequenza di scelte corrette. Dopo un paio di minuti, però, i punteggi calarono in modo drastico, raggiunsero un minimo a una distanza di tempo compresa fra i due e tre minuti e poi salirono di nuovo al livello iniziale.

Questo genere di relazione fra formazione del ricordo e tempo appare anche in altri animali e nell'uomo. I meccanismi fisiologici sottostanti all'associazione fra

Un esperimento con i colori eseguito da Karl von Frisch nel secondo decennio del secolo accertò la capacità delle api di associare colori al cibo e di ricordare un colore così appreso. Egli mise un piatto contenente una soluzione di miele su un tavolo. Le api ne furono attratte. Collocò poi il piatto su un pezzo di cartone blu, così che il blu era il colore che le api vedevano quando si avvicinavano al piatto e succhiavano il miele. Von Frisch dispose poi attorno al pezzo di cartone blu altri pezzi di cartone aventi la medesima figura ma di colore bianco, nero e in varie tonalità di grigio. Su ogni cartone, compreso quello blu, posò un piatto vuoto. Le api erano state condizionate ad attendersi di trovare cibo sul tavolo e perciò continuarono a venire. La maggior parte di esse andavano a vedere solo nel piatto che era posto sul cartone blu.

un segnale e una ricompensa (o punizione) hanno bisogno di tempo (secondi, minuti e anche ore) e molto probabilmente passano per fasi in cui il sostrato per la fissazione del ricordo è diverso. Nel caso tipico il ricordo viene fissato dapprima in una memoria sensoriale, da cui l'informazione può svanire rapidamente se non viene rinforzata. Attraverso il rinforzo, l'informazione passa a una memoria a breve termine, la quale si riflette nell'alto livello di risposte corrette date da un'ape immediatamente dopo una o più gratificazioni. Questo aumento, dopo un breve intervallo di declino, rappresenta una fase designata come consolidamento, in cui l'informazione viene trasferita nella memoria a lungo termine.

Nel caso di un'ape, come ci si potrebbe attendere, le fasi di apprendimento sono adattate molto bene alla situazione di apprendimento quale si presenta in natura. I fiori per la maggior parte offrono quantità di nettare così piccole che un'ape impiega di solito solo pochi secondi a esaurirne la scorta; pochi secondi dopo si posa su un altro fiore. Se i fiori sono sparsi a una certa distanza l'uno dall'altro, l'ape può impiegare anche un minuto prima di trovare un altro fiore simile. La frequenza inizialmente alta di scelte corrette indica un meccanismo a breve termine tale da garantire che l'ape troverà un fiore simile anche dopo gratificazioni brevissime. La fase di consolidamento più a lungo termine sembra essere adattata al volo dell'ape fra l'area di raccolta del nettare e l'alveare. Quando l'ape ritorna nel prato dall'alveare, compie di nuovo un alto livello di scelte corrette, indicando così che l'informazione è stata fissata nella memoria a lungo termine.

Per quanto tempo un segnale cromatico rimane fissato nella memoria sensoriale dell'ape? La risposta consiste nell'accertare le condizioni alle quali un colore continuerà a essere associato con del cibo anche quando il colore non è più visibile. Nei nostri test in proposito abbiamo stabilito che un segnale cromatico percepito da un'ape fino a due secondi prima dell'inizio della gratificazione viene appreso altrettanto bene che se il segnale fosse presente per l'intero periodo in cui l'ape si accosta alla fonte di cibo e lo sugge. Se un segnale cromatico viene presentato dopo che l'ape ha iniziato a succhiare la soluzione, esso non viene associato al cibo.

Elisabeth Opfinger osservò nel 1931, quando studiava sotto la guida di von Frisch, che l'ape apprende un colore solo se lo vede mentre si avvicina al fiore, mentre non impara i colori che vede mentre sta succhiando e allontanandosi dal fiore. Oggi sappiamo che un segnale cromatico, per essere appreso, dev'essere presente non più di tre secondi prima e non più di mezzo secondo dopo l'istante in cui l'ape inizia a suggere il nettare. Poiché un'ape

Le curve di apprendimento indicano la rapidità con cui le api apprendono i vari colori. Le curve rappresentano valori medi ricavati dal comportamento di molte api. La media delle risposte corrette aumenta col numero delle gratificazioni (offerte di cibo). In questi esperimenti, in alternativa al colore premiato era usato il colore complementare: per l'ultravioletto a 361 nanometri era il blu-verde a 494; per il violetto a 428 nanometri era l'arancione a 590 e per il blu era il «violetto delle api», un colore misto formato per l'80 per cento da arancione e per il 20 per cento da ultravioletto *(curva più a destra)*. L'apprendimento è particolarmente rapido per il violetto a 408 e 418 nanometri, mentre la massima lentezza si riscontra nell'apprendimento del blu-verde a 494 nanometri. Addestrate al riconoscimento del violetto, le api raggiungono un livello di precisione del 90 per cento dopo una gratificazione; per il blu-verde hanno bisogno invece di circa cinque gratificazioni.

Il condizionamento delle api all'odore offre un altro modo per sottoporre a test la formazione della memoria. L'ape viene immobilizzata in un tubo e una delle sue antenne è toccata con una goccia di soluzione zuccherina. L'ape estende la proboscide per riflesso incondizionato. Se, prima di toccare l'antenna con la soluzione zuccherina, si presenta all'ape un soffio di odore, e le si consente di sentire il sapore dello zucchero con la proboscide, l'odore diventa uno stimolo condizionante. Dopo un solo condizionamento (un assaggio di saccarosio), per la maggior parte le api estendono la proboscide anche quando viene presentato loro solo il soffio di odore.

momento che il blocco dell'attività neurale sia nelle api addestrate al riconoscimento di colori, che venivano lasciate volare liberamente, sia nelle api addestrate al riconoscimento di odori, alle quali impedimmo di volare, dette risultati simili nel corso del tempo, pare che i processi nervosi centrali che sono alla base dell'immagazzinamento di informazioni nella memoria precoce siano simili nei due casi dell'apprendimento del colore e dell'odore.

Abbiamo identificato aree particolari del cervello che partecipano alla formazione della memoria degli odori. Se queste medesime aree siano correlate anche all'apprendimento visivo è una questione ancora aperta. In ogni caso i decorsi temporali dell'apprendimento dell'informazione tanto olfattiva quanto visiva sono del tutto simili. Registrazioni elettrofisiologiche eseguite sulle nostre api immobilizzate e condizionate al riconoscimento di odori dovrebbero chiarire perciò anche i meccanismi generali che operano nelle api lasciate volare liberamente.

Non è facile fare registrazioni dell'attività fisiologica di singoli neuroni nel cervello delle api poiché i neuroni sono di solito molto piccoli. Siamo riusciti nondimeno a eseguire tali registrazioni in api immobilizzate, nel corso di esperimenti nei quali l'insetto era libero di muovere la sua proboscide e poteva essere condizionato all'odore. I nostri studi di elettrofisiologia sono soltanto agli inizi ma disponiamo già di alcune scoperte.

La maggior parte dei neuroni presenti nel corpo peduncolato riceve informazioni da più di un sistema sensoriale. Un neurone, per esempio, potrebbe rispondere a un lampo di luce con un'inibizione, a una soluzione zuccherina portata a contatto con una sua antenna con uno stato di eccitazione e, a un odore misto, ancora con uno stato di eccitazione. Abbiamo osservato molte combinazioni diverse di risposte a stimoli diversi.

Solo un piccolo numero di neuroni modifica la sua sensibilità durante il condizionamento. Gli unici che modificarono la loro sensibilità erano già sensibili all'odore e allo zucchero prima del condizionamento. Il mutamento consiste di solito in un aumento della frequenza di scarica spontanea del neurone, accompagnato a un aumento della sensibilità a stimoli olfattivi e alla soluzione zuccherina. In alcuni neuroni trovammo una stretta correlazione fra il numero di condizionamenti e il mutamento nella sensibilità. Attraverso una colorazione intercellulare di neuroni, stiamo tentando di identificare le particolari cellule che apprendono. La maggior parte di queste risultano localizzate nell'area circostante i corpi peduncolati, corroborando così l'ipotesi che questa parte del cervello sia importante per l'elaborazione e l'immagazzinamento dell'informazione olfattiva.

Dal lavoro svolto da noi e da altri sulle api risulta evidente che nelle api l'apprendimento associativo procede secondo le fasi principali dell'apprendimento che sono state accertate per i vertebrati. Le api hanno inoltre una memoria sensoriale a breve termine e una a lungo termine del tutto simili a quelle che si trovano in varie specie di vertebrati. Lavorando sulle api dovremmo avere perciò buone probabilità di individuare alcune caratteristiche fondamentali dell'apprendimento associativo e della memoria che siano valide anche per i complessi sistemi nervosi dei vertebrati.

Rappresentazione schematica del cervello di ape in vista frontale. L'informazione visiva viene elaborata principalmente dalla lamina, dalla medulla e dai lobuli. Il mesencefalo comprende due corpi fungiformi, i corpi peduncolati, costituiti da neuroni molto compatti e intrecciati. Nei corpi peduncolati si distinguono i calici e i lobi alfa e beta. Altre due strutture, i lobi antennali, sono i centri sensoriali primari per le informazioni provenienti dalle antenne. Le vie nervose principali che elaborano il flusso d'informazione fra i corpi peduncolati e i lobi antennali sono in colore. Il cervello delle api contiene circa 850 000 cellule nervose in un volume di un millimetro cubo.

Applicazioni industriali delle membrane sintetiche

Sottili pellicole di polimeri sono in grado di separare molecole in base a dimensioni, carica o altre proprietà. Questo le rende particolarmente adatte a vari impieghi quali il trattamento dei rifiuti o la dissalazione

di Harry P. Gregor e Charles D. Gregor

Un sorprendente numero di processi industriali può essere semplicemente classificato come procedimento di separazione di una sostanza da un'altra. Nell'attività mineraria, per esempio, occorre separare metallo puro dal minerale in cui è contenuto. La separazione del sale da cucina dall'acqua di mare fornisce due prodotti utili: lo stesso sale e acqua potabile. Nel trattamento delle acque luride e di molti effluenti industriali, i rifiuti, per poter essere efficacemente eliminati, devono venire concentrati separandoli dal fluido in cui scorrono. Un modo per svolgere direttamente processi come quelli sopra descritti potrebbe consistere nella realizzazione di una barriera in grado di consentire il passaggio di particelle o molecole di una certa specie e di impedire quello di tutte le altre. È possibile costruire una simile barriera selettiva sotto forma di una membrana sintetica di natura polimerica. Per la verità, da molti anni si conoscono membrane di questo genere, però soltanto in epoca recente sono stati superati alcuni ostacoli che ne impedivano l'impiego commerciale.

Le membrane sintetiche sono in grado di separare le diverse sostanze sulla base di numerose proprietà. Nel processo denominato ultrafiltrazione, le molecole vengono separate in base alle loro dimensioni, nell'elettrodialisi in base alla loro carica elettrica, nell'estrazione con solvente in base alle loro diverse affinità per due solventi di natura dissimile. Nel processo chiamato osmosi inversa, si fa uso di membrane che sono permeabili quasi esclusivamente all'acqua. Disponendo opportunamente numerose membrane, è possibile decomporre una soluzione acquosa di un sale in un acido e in una base e, infine, associando un enzima a una membrana, possono essere attivate le reazioni chimiche aventi come effetto un'efficace separazione tra reagenti e prodotti.

Forse, l'uso tecnologico più diretto di membrane si ha nell'ultrafiltrazione, in cui molecole disciolte e particelle sospese vengono vagliate essenzialmente in base al fatto se siano o meno sufficientemente piccole da attraversare i pori di una membrana. Nella normale filtrazione, un liquido viene forzato a passare sotto pressione attraverso un mezzo poroso, che può essere un foglio di carta o un letto di sabbia compatta. L'ultrafiltrazione non differisce sostanzialmente da questo schema, se si eccettuano le dimensioni dei pori, che sono circa 1000 volte più piccoli di quelli di un filtro normale, circostanza questa che rende possibile la separazione e la concentrazione non solo di materiale particellare, ma anche di molecole prestabilite. Alcune membrane per ultrafiltrazione tendono, inoltre, a escludere molecole elettricamente cariche, indipendentemente dalle loro dimensioni.

Quasi tutte le particelle sospese hanno dimensioni dell'ordine di 10^{-6} metri, cioè di un micrometro. Le particelle di polvere raramente hanno diametro inferiore al micrometro, mentre i batteri raggiungono al massimo la lunghezza di 5-10 micrometri. I pori di una membrana per ultrafiltrazione sono dell'ordine di grandezza di 10^{-9} metri, ossia di un nanometro; i diametri medi dei pori variano da meno di un nanometro a circa 10 nanometri. Di conseguenza, pressoché tutte le particelle sospese, compresi i batteri, sono separabili per ultrafiltrazione, come pure i virus che hanno una dimensione minima di 20 nanometri.

L'ultrafiltrazione è, nella sua essenza, una filtrazione molecolare, capace di distinguere tra molecole di differenti dimensioni e, grossolanamente, tra molecole di diverso peso molecolare. L'acqua, con un peso molecolare di 18 e con un diametro effettivo di 0,2 nanometri, può attraversare liberamente tutte le membrane per ultrafiltrazione. Anche gli zuccheri semplici, come il glucosio (con peso molecolare di 180 e una dimensione di circa 1,5 nanometri), possono attraversare la maggior parte delle membrane per ultrafiltrazione. Le proteine e altre grandi molecole biologiche sono invece generalmente arrestate, poiché spesso hanno peso molecolare che si avvicina a 100 000 e dimensioni di parecchi nanometri. Scegliendo una membrana con pori di dimensioni attentamente controllate è anche possibile separare molecole con limitate differenze di peso molecolare, come il saccarosio (peso molecolare 342) e la cianocobalamina o vitamina B 12 (peso molecolare 1300).

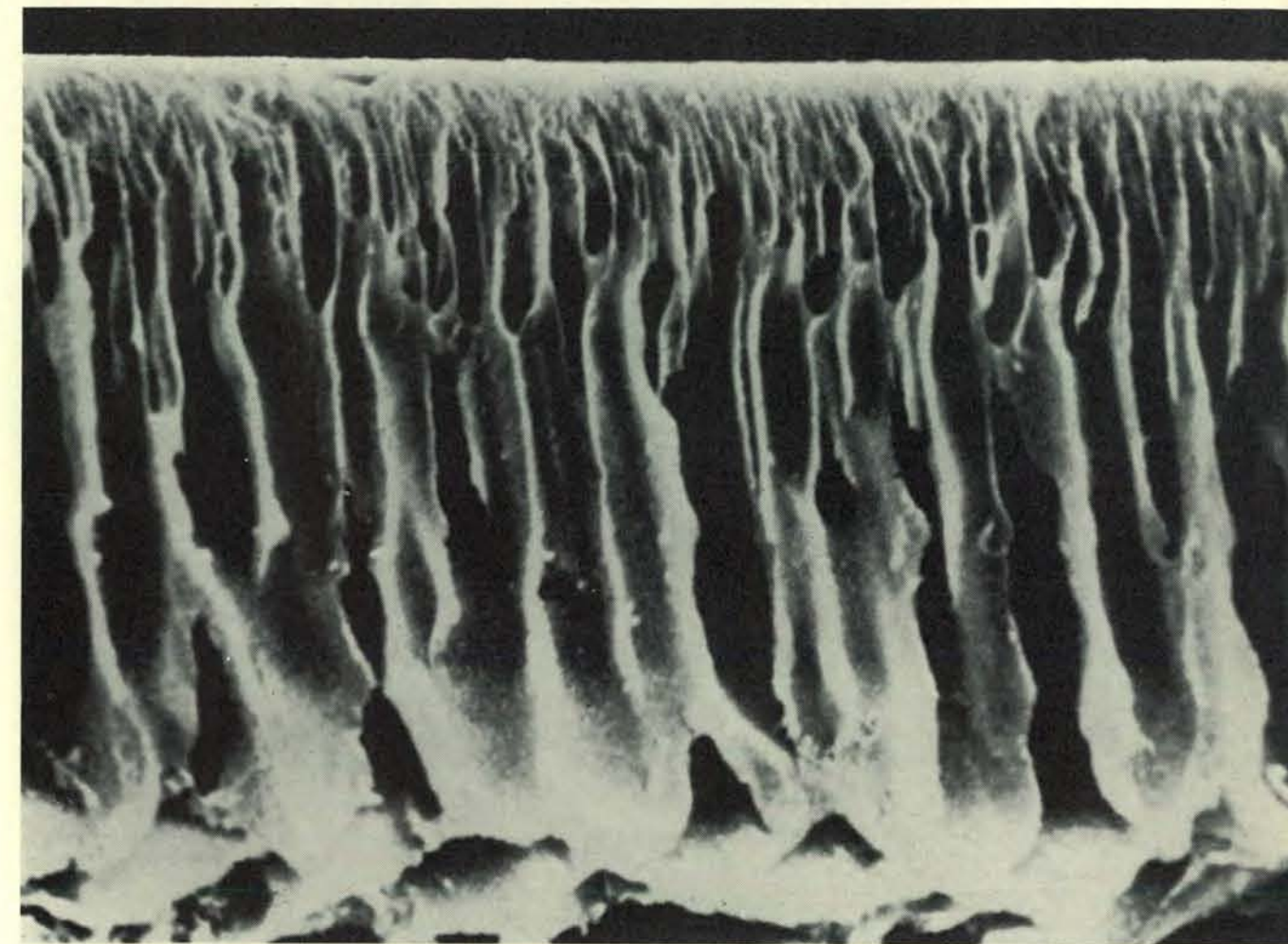
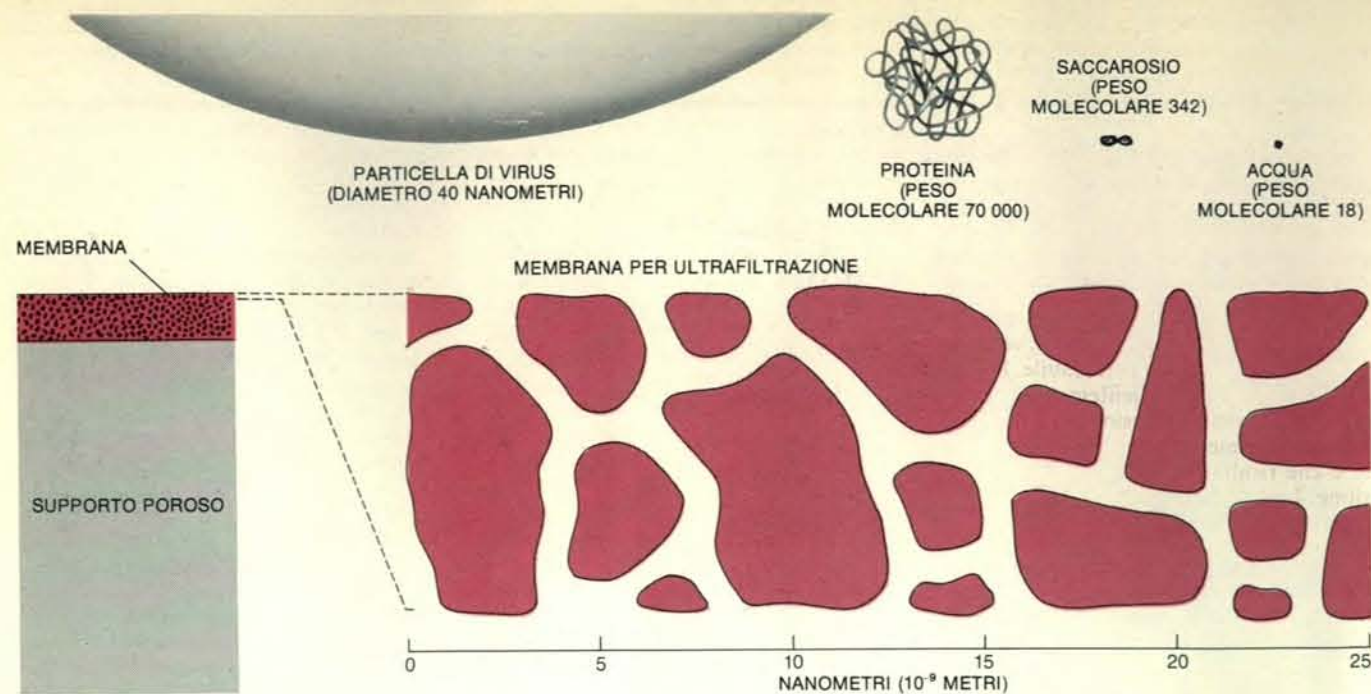
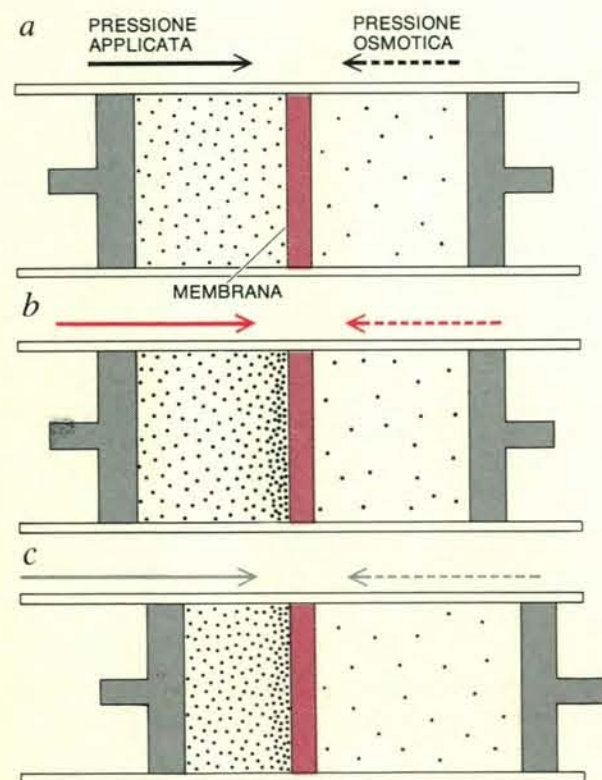
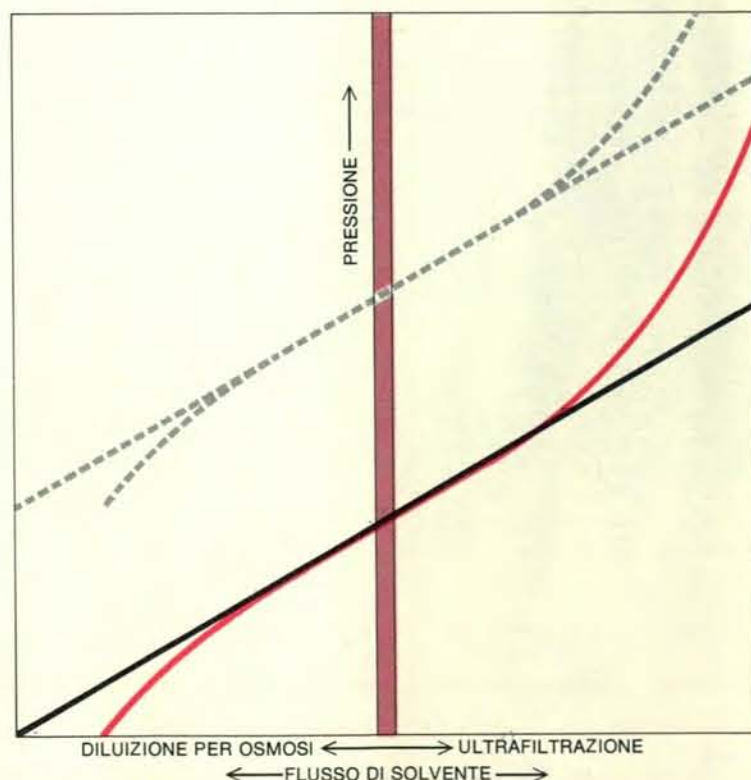
Una membrana per ultrafiltrazione viene ottenuta da una soluzione di polimeri a catena lunga, come per esempio l'acetato di cellulosa sciolto in acetone. La soluzione è viscosa e viene distesa in modo da formare uno strato sottile. Viene quindi lasciato evaporare parte del solvente e, come risultato, il polimero comincia a separarsi dalla soluzione sotto forma di una sottile matrice semisolida. La membrana viene infine immersa in un altro solvente, per esempio acqua, che fa precipitare rapidamente tutto il rimanente polimero. È questo raffreddamento che dà origine ai pori della membrana, in quanto la rapida precipitazione determina il rapprendimento del polimero. Le dimensioni dei pori sono in larga misura determinate dalla durata dell'evaporazione parziale e dalla scelta dei solventi. Sebbene i pori non risultino mai completamente uniformi, è possibile controllarne il diametro medio in modo piuttosto rigoroso.

Le membrane ottenute in questo modo sono assai sottili, generalmente da cinque a 30 micrometri e, poiché sono chiamate a sopportare una pressione, è necessario provvederle di un supporto meccanico, per mezzo di un substrato avente porosità macroscopica. È possibile, dopo aver ottenuto la membrana su una lastra di vetro, staccarla e montarla su un substrato idoneo. La tecnica cui si fa attualmente più comune ricorso consiste nel «gettare» direttamente la membrana su un supporto poroso, che può essere plastica, carta o tessuto.

Il primo dispositivo per ultrafiltrazione con membrana sintetica fu probabilmente quello realizzato dal chimico e biologo tedesco Moritz Traube nel 1870. Traube precipitò un gel di ferrocianuro rameico nei pori macroscopici da un vaso da fiori non verniciato. Negli anni trenta, furono prodotte membrane polimeriche per ultrafiltrazione con pori di differenti dimensioni. Riveste particolare importanza il lavoro svolto a quell'epoca da William J. Elford che, oltre a ottenere membrane di acetato e di nitrato di cellulosa, chiarì la maggior parte dei principi del processo di ultrafiltrazione.

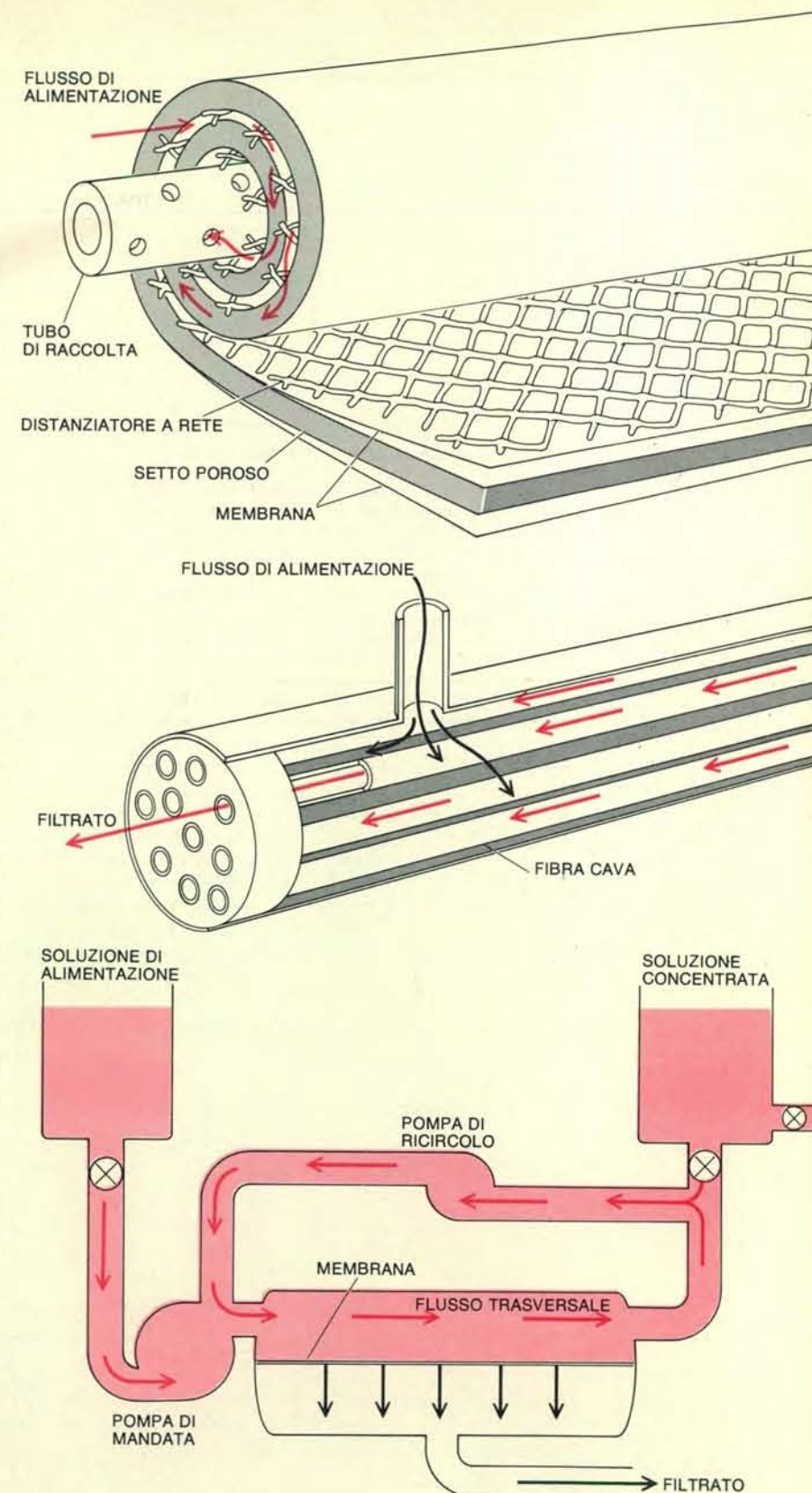
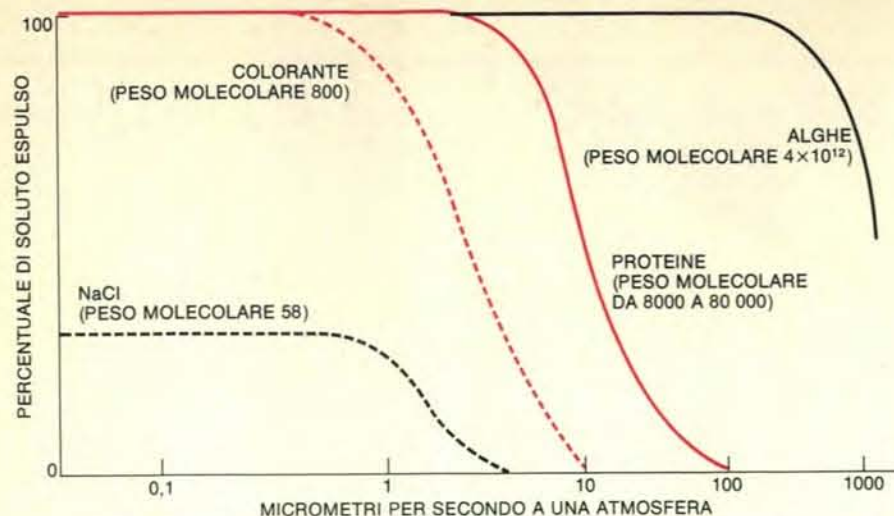
La teoria dell'ultrafiltrazione può essere illustrata sulla scorta di un semplice modello costituito da una membrana interposta tra due liquidi, uno dei quali è acqua pura, e l'altro acqua in cui è disciolta una qualche sostanza. In prima istanza, si può partire dall'ipotesi che la membrana sia assolutamente permeabile all'acqua e completamente impermeabile al soluto. L'entità e la direzione del flusso attraverso la membrana vengono, di conseguenza, regolate da una combinazione di quattro fattori.

Il primo di questi fattori è una pressione osmotica che tende a sospingere acqua, dalla parte della membrana in contatto con acqua pura, verso la soluzione, attraverso la membrana stessa. Una rigorosa spiegazione dell'origine della pressione osmotica può essere svolta solo introducendo le leggi della termodinamica e, in particolare, il concetto di entropia; tuttavia, si possono afferrare intuitivamente le ragioni di questo fenomeno se lo si considera come manifestazione della tendenza del solvente (in questo caso acqua) a uniformare la propria concentrazione attraverso l'intero volume del liquido, come se, in un certo senso, la membrana non fosse presente. Poiché la concentrazione dell'acqua pura è maggiore di quella dell'acqua in soluzione, le sue molecole tendono a trasferirsi nella soluzione passando attraverso la membrana. L'entità della pressione osmotica corrisponde alla differenza nella concentrazione dell'acqua da entrambi i lati della membrana.

Se il sistema della membrana fosse lasciato a se stesso, l'acqua pura fluirebbe in modo da diluire la soluzione, che è l'effetto opposto a quello desiderato. È quindi necessario vincere la pressione osmotica applicando una contropressione alla soluzione. Se quest'ultima è maggiore della pressione osmotica, si avrà un deflusso di acqua dalla soluzione, con conseguente aumento della concentrazione del soluto. La velocità di flusso è proporzionale alla differenza tra le due pressioni ed è inoltre influenzata da altri due fattori: è direttamente proporzionale alla permeabilità all'acqua della membrana e inversamente proporzionale allo spessore della mem-

Microfotografia eseguita al microscopio elettronico della sezione trasversale di una membrana porosa (l'ingrandimento è di circa 500 volte). La membrana è stata ottenuta a partire da una soluzione di polimeri che, dopo essere stata stesa in strato sottile e fatta essiccare parzialmente, è stata immersa in acqua. In seguito alla rapida estrazione del solvente residuo e al rapprendimento del polimero a opera dell'acqua, si sono formati i pori che appaiono nella microfotografia come canalicoli verticali. I pori più fini vicino alla superficie superiore della membrana si sono prodotti in quanto gli strati superficiali si sono essiccati più completamente, in modo da formare una «buccia» sulla soluzione di polimero. La membrana, fabbricata dalla Amicon Corporation, è stata studiata per essere impiegata nell'ultrafiltrazione, processo attraverso il quale è possibile allontanare da un fluido non soltanto particelle sospese, ma anche molte molecole presenti in soluzione.

Una membrana per ultrafiltrazione possiede pori di dimensioni molecolari. La membrana stessa è un film avente spessore da 5 a 30 micrometri, depositato su un substrato poroso molto più spesso, quale carta o tessuto, che funge da supporto meccanico. Si possono realizzare membrane con pori aventi un diametro medio compreso tra 1 e 10 nanometri. La membrana qui schematicamente rappresentata in sezione trasversale *(in basso a destra)* possiede pori del diametro di 1 nanometro. L'acqua attraversa agevolmente i pori, al pari di molte piccole molecole, quali il saccarosio. La maggior parte delle proteine viene invece respinta, così come i virus di maggiori dimensioni e altri microrganismi. Una membrana per ultrafiltrazione con pori di opportune dimensioni può operare separazioni tra molecole come zuccheri e proteine.

Il flusso di acqua o di altro solvente attraverso una membrana per ultrafiltrazione dipende in parte dalle proprietà intrinseche della membrana e in parte dalla pressione. L'esistenza di una differenza di concentrazione genera una pressione osmotica, che tende a sospingere acqua pura, attraverso la membrana, nella soluzione più concentrata, instaurando un flusso nella direzione opposta a quella desiderata. Se si deve effettuare una filtrazione, questa retropressione osmotica deve essere superata applicando alla soluzione una pressione di maggiore entità. Nel caso ideale *(a, curva in nero)*, il flusso attraverso la membrana è semplicemente proporzionale alla differenza tra la pressione applicata e la pressione osmotica. In pratica, in ogni caso in cui sussista un flusso di entità apprezzabile, si determina un accumulo di molecole di soluto lungo la superficie della membrana prospiciente la soluzione concentrata *(b, curva in colore)*; questa condizione, chiamata polarizzazione di concentrazione, aumenta la retropressione osmotica e richiede un aumento nella pressione applicata se si vuole mantenere un dato flusso. Inoltre, se la soluzione di ingresso non viene continuamente diluita, la sua concentrazione aumenta con la filtrazione *(c, curva in grigio)*.

brana. Poiché anche una soluzione moderatamente concentrata è in grado di generare una pressione osmotica di alcune decine di chilogrammi per centimetro quadrato, nell'ultrafiltrazione è spesso necessario applicare pressioni elevate.

Nell'ultrafiltrazione, una molecola che risulti troppo voluminosa per attraversare i pori di una membrana viene con assoluta certezza respinta, ma possono anche venire concentrate, in una certa misura, molecole di dimensioni inferiori, verso le quali la membrana sia permeabile. Il flusso di tali soluti è prevalentemente regolato dal processo di diffusione, che rappresenta il movimento casuale delle molecole e che risulta indipendente dalla pressione. Le molecole di un soluto capaci di attraversare una membrana diffondono attraverso di essa in entrambe le direzioni, ma la risultante netta del movimento è verso la regione di minore concentrazione. L'entità del flusso è direttamente proporzionale alla differenza nelle concentrazioni e inversamente proporzionale allo spessore della membrana. Poiché anche il movimento del solvente è inversamente proporzionale allo spessore della membrana, quest'ultimo non ha influenza nell'impedire di fatto il passaggio di un soluto capace di attraversare la membrana. Tale impedimento varia solo in funzione della pressione applicata in quanto, al crescere della pressione, aumenta il flusso di acqua, senza che venga modificato il flusso del soluto.

Alla base di questa descrizione del fenomeno sta l'ipotesi che la concentrazione dell'acqua e del soluto siano costanti ai due lati della membrana. Nella realtà, questa condizione non può essere soddisfatta. Immediatamente adiacente alla membrana esiste sempre uno strato limite di soluzione stagnante, dove si verifica una continua estrazione di acqua e un accumulo, fino a raggiungere concentrazioni insolitamente elevate, di soluti impediti nel loro passaggio. Questo fenomeno, detto polarizzazione di concentrazione, aumenta la pressione osmotica che si oppone alla filtrazione, poiché la differenza nelle concentrazioni del solvente che determina l'effetto osmotico è quella dei fluidi più vicini alla membrana. L'entità della polarizzazione di concentrazione aumenta proporzionalmente al flusso di solvente ed è quindi proporzionale anche rispetto alla pressione applicata. Conseguentemente sussiste, per ogni membrana, una velocità massima di flusso per unità di superficie, oltre la quale ogni ulteriore aumento della pressione applicata ha come effetto, anziché una intensificazione del flusso, un aumento della retropressione osmotica. La polarizzazione di concentrazione può anche dare origine ad altri fenomeni indesiderati, quali per esempio la precipitazione di solidi sulla superficie della membrana. Per impedire questo inconveniente bisogna limitare la pressione applicata.

Il flusso attraverso una membrana per ultrafiltrazione può essere misurato in termini di volume trasferito per unità di

Un'apparecchiatura per ultrafiltrazione deve essere studiata in modo che venga ridotto al minimo il fenomeno dell'otturazione dei pori e venga attenuata la polarizzazione di concentrazione. In una disposizione detta «a rotolo di gelatina» *(in alto)*, due membrane separate da un setto poroso sono avvolte, insieme a un distanziatore di plastica a forma di rete, attorno a un tubo poroso. La soluzione viene pompata attraverso il distanziatore e il liquido filtrato, dopo aver attraversato la membrana, si muove a spirale attraverso il setto fino al tubo al centro, dove viene raccolto. Si possono anche «filare» membrane sintetiche sotto forma di sottili fibre cave *(al centro)* con un diametro interno di circa 40 micrometri. Le fibre sono disposte a fascio e la soluzione di ingresso viene pompata intorno a esse, mentre il liquido filtrato fuoriesce dalle loro estremità. Ogni cella per ultrafiltrazione è munita di una pompa che fornisce la pressione necessaria al processo.

La permeabilità delle membrane a molecole di diverse dimensioni è legata alle dimensioni dei pori, che vengono a loro volta misurate dalla velocità alla quale l'acqua fluisce attraverso la membrana a una pressione applicata convenzionale. Le membrane più rigide permettono il passaggio di significative quantità di cloruro di sodio, quelle più permeabili arrestano le cellule algali. Tra questi estremi, vi sono talune membrane che respingono solo molecole di grandi dimensioni, quali quelle proteiche, e altre che bloccano molecole moderatamente voluminose.

superficie di membrana per effetto di una determinata differenza di pressione. Come ci si può attendere, il flusso aumenta al crescere delle dimensioni dei pori. L'unità è espressa, nel sistema metrico decimale, in micrometri per secondo, a una differenza di pressione pari a una atmosfera.

Poiché la misurazione diretta delle dimensioni dei pori in una membrana sarebbe alquanto difficoltosa, si può ricorrere a una loro stima misurando la velocità di flusso in condizioni convenzionali che è, entro certi margini, determinata dalle dimensioni dei pori. Il flusso misurato può essere calibrato osservando le modalità con cui vengono respinte molecole di varie dimensioni da una membrana di cui sia nota la velocità di flusso. Come tipiche molecole di prova si potrebbero usare l'albumina del siero di bue (una proteina avente peso molecolare 75 000), il colorante eritrosina (peso molecolare 800) e il saccarosio (peso molecolare 342). Una membrana con una velocità di flusso di appena un micrometro al secondo a un'atmosfera respinge quasi completamente le molecole più voluminose e trattiene una parte dello zucchero. Una membrana con una velocità di flusso medio, di circa tre micrometri al secondo a un'atmosfera, respinge il colorante e la proteina, ma lascia passare la maggior parte dello zucchero. Una membrana definibile come «rapida», con una velocità di flusso da 10 a 20 micrometri al secondo, respinge solo la proteina, mentre una membrana assai rapida (con una velocità di flusso superiore a 50 micrometri al secondo) lascia passare anche la proteina, mentre esclude materiale particellare come microrganismi.

Nella normale filtrazione, il liquido di alimentazione, col suo materiale sospeso e disciolto, viene semplicemente pompato attraverso il mezzo filtrante, che è spesso un letto di sabbia compatta. Le particelle finemente disperse nel liquido aderiscono al filtro e le une alle altre, formando uno strato di sporco, che è noto con il termine tedesco *Schmutzdecke*.

Poiché lo schmutzdecke diventa presto impermeabile, il filtro deve essere periodicamente sostituito o lavato in controcorrente per mantenere una soddisfacente velocità di flusso. Nell'ultrafiltrazione viene impiegata una tecnica differente. La soluzione di alimentazione viene fatta continuamente circolare attraverso la superficie della membrana e, con questo accorgimento, si contribuisce non solo a evitare la formazione di depositi otturanti, ma anche a ridurre la polarizzazione di concentrazione.

I componenti dell'apparecchiatura di ultrafiltrazione possono essere variamente disposti. Si possono, per esempio, montare su un telaio fogli paralleli di membrana, oppure quest'ultima può rivestire internamente un tubo poroso. Una disposizione particolarmente efficace è quella a «rotolo di gelatina», ottenuta avvolgendo membrane e distanziatori attorno a un tubo poroso che funge da raccoglitore del liquido filtrato. La membrana viene saldata su ciascun lato di un foglio di carta porosa o di tessuto e quindi la doppia membrana viene avvolta insieme a un distanziatore a rete in plastica, sul tubo di raccolta. La soluzione di alimentazione viene pompata attraverso gli strati della doppia membrana e il filtrato si muove a spirale attraverso il mezzo poroso fino a raggiungere il tubo al centro. Questo dispositivo è sostanzialmente più economico dei filtri di membrana montati su un telaio o di quelli formati da un tubo rivestito internamente dalla membrana e assicura una forte corrente trasversale lungo la superficie della membrana sul lato dove si concentrano i soluti.

È anche possibile ottenere una membrana sotto forma di fibre cave, generalmente non più spesse di un capello umano. Poiché le fibre hanno pareti spesse se paragonate alle loro dimensioni complessive, possono, di regola, sopportare pressioni dell'ordine della tonnellata per centimetro quadrato. Il costo è circa lo stesso del dispositivo avvolto a spirale, ma la conformazione della fibra cava la rende suscettibile di otturarsi. Se il liquido di alimentazione viene fatto passare attraverso l'anima cava, le fibre tendono a otturarsi di particelle mentre, se viene fatto circolare intorno alla parete esterna delle fibre, il flusso trasversale è inadeguato.

EPARINA

ACIDO POLISTIRENESOLFONICO

ACIDO POLIVINILSOLFONICO

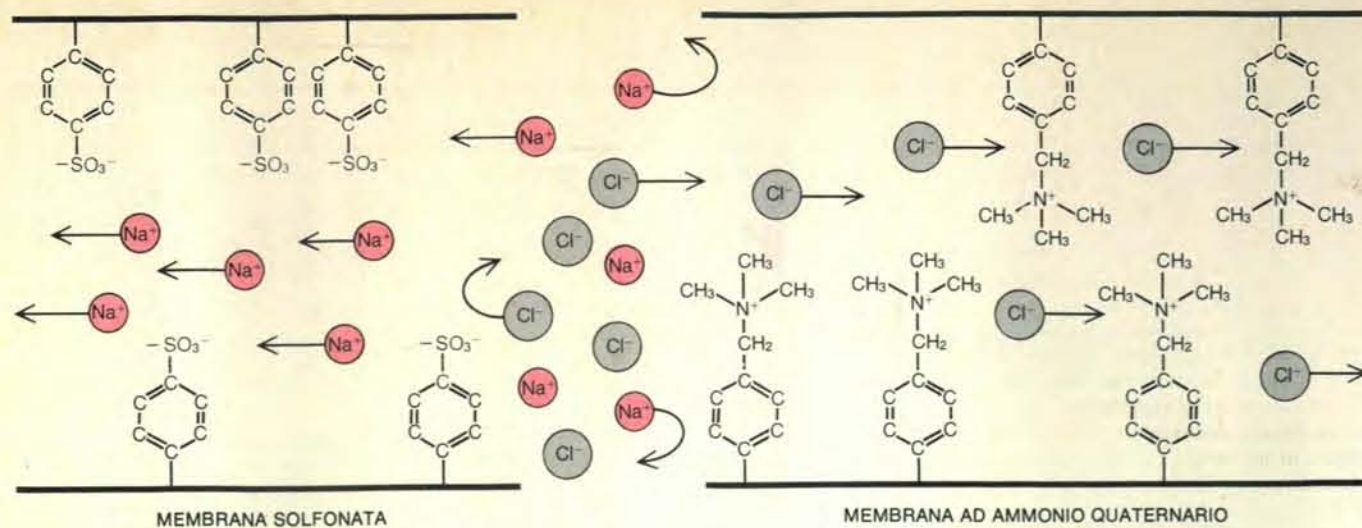
I polimeri «non otturabili» incorporano nella propria struttura il gruppo solfonico ($-SO_3$), che conferisce loro una elevatissima affinità per l'acqua, impedendo quindi alla maggior parte del materiale estraneo di aderire. La maggior parte delle particelle otturanti sono idrofobe e in prevalenza trasportano una carica negativa. L'eparina è un agente anticoagulante biologico, che mantiene «bagnati» i vasi sanguigni. I gruppi chimici attivi dell'eparina sono anch'essi solfonici.

L'osmosi inversa è simile all'ultrafiltrazione in quanto viene forzata una soluzione acquosa sotto pressione attraverso una membrana e, mentre l'acqua passa, i soluti sono bloccati. Il meccanismo è tuttavia completamente diverso. La membrana per osmosi inversa non possiede pori definiti, bensì interstizi tra le fibre di un polimero dove può trovare posto un piccolo volume di acqua. Negli interstizi l'acqua assume una struttura simile a quella del ghiaccio (*in colore*), in cui le molecole presentano una disposizione ordinata, oltre a essere collegate da legami a idrogeno. (Anche nell'acqua pura esistono piccoli raggruppamenti a struttura simile al ghiaccio, in uno stato di continua formazione e disintegrazione.) La struttura simile al ghiaccio ha una simmetria tetraedrica in quanto ciascun atomo di ossigeno è circondato da altri quattro equidistanti (qui ne viene data una rappresentazione proiettata su un piano). Le molecole d'acqua sotto pressione entrano a far parte della configurazione simile al ghiaccio dal lato della membrana prospiciente la soluzione concentrata (*a sinistra*) e fuoriescono per fusione dall'altro lato (*a destra*). Le altre molecole e particelle vengono respinte, non solo quelle troppo voluminose per adattarsi alla membrana, ma anche quelle così piccole da non conformarsi alla struttura simile a ghiaccio. Gli ioni, in particolare, sono arrestati perché il rivestimento di molecole d'acqua che li circonda potrebbe attraversare il reticolo cristallino simile a ghiaccio solo distruggendolo, cosa che non accade.

L'ultrafiltrazione, pur essendo stata sviluppata negli anni venti, non ha trovato per altri 40 anni impieghi, se non rari, al di fuori della pratica di laboratorio, principalmente a causa di uno specifico problema: l'otturazione delle membrane. Ogni qualvolta una membrana veniva esposta a un flusso di alimentazione contenente solidi sospesi, le particelle aderivano alla superficie ostruendone i pori. Ciò determinava un rapido rallentamento nella velocità di flusso e, nello spazio di minuti o di ore, il filtro diventava inutilizzabile.

Intorno al 1960, uno di noi (Harry P. Gregor) iniziò uno studio sistematico del meccanismo di otturazione delle membrane, con l'obiettivo di realizzare membrane non suscettibili di questo fenomeno. Gli inconvenienti più gravi erano dovuti a materiali quali argille, particelle oleose e proteine, aventi tutti in comune estese superfici con caratteristiche idrofobe (che respingono l'acqua). Quando una sostanza idrofoba si trova in un mezzo acquoso, può ridurre la propria energia complessiva diminuendo la superficie esposta all'acqua; due particelle idrofobe tendono, per esempio, ad agglomerarsi, espellendo l'acqua contenuta nello spazio tra l'una e l'altra e riducendo pertanto la propria superficie esposta. Analogamente, una simile particella può essere trattenuta alla superficie di una membrana in seguito alla neutralizzazione delle interazioni repulsive con l'acqua circostante. La maggior parte dei materiali che producono depositi otturanti trasportano anche una carica elettrica negativa e i legami a idrogeno che si instaurano per la presenza di queste cariche potrebbero dare un ulteriore contributo alla formazione di depositi. In questo tipo di legame, la debole carica positiva di un atomo di idrogeno sulla superficie della membrana attrae un gruppo carico negativamente presente nella particella otturante.

Un metodo abbastanza ovvio per evitare le interazioni di tipo idrofobo è quello di realizzare una membrana che sia fortemente idrofila, cioè che possegga una assai spiccata affinità per l'acqua. Un materiale di questo tipo è in grado di rimanere umido anche in presenza di particelle idrofobe, impedendo quindi alle particelle di aderire alla propria superficie attraverso l'espulsione di acqua.

Tra i polimeri conosciuti, i più idrofili sono quelli recanti alla superficie il gruppo solfonico $-SO_3^-$ (Un acido solfonico ha come formula generale $R-SO_3H$ in cui R rappresenta un radicale alifatico o aromatico). I polimeri solfonati manifestano una eccezionale affinità per l'acqua, al pari dell'acido solforico. Ciascun gruppo solfonico è comunemente circondato da molte molecole d'acqua, che non possono essere estromesse da particelle idrofobe. Il gruppo $-SO_3^-$ non è inoltre suscettibile di dare luogo a legami a idrogeno, in quanto la sua carica negativa respinge le particelle otturanti cariche negativamente. La carica negativa del gruppo solfonico si conserva anche in presenza di una elevata concentrazione di ioni idrogeno (cioè, a pH bassi), poiché il gruppo è la base coniugata di un acido molto forte. Il gruppo solfonico resiste inoltre alla disattivazione a opera di ioni positivi, come quelli dei metalli pesanti. Anche i sali di piombo e di bario dei polimeri solfonati rimangono perfettamente solubili in acqua, e ciò sta a indicare che i gruppi solfonici sono ancora idratati, mentre la mag-

Le membrane cariche sono in grado di discriminare gli ioni in soluzione in base alla loro carica elettrica. Sono qui rappresentati i singoli pori di due membrane cariche. La membrana carica negativamente *(a sinistra)* possiede gruppi solfonici lungo le pareti dei pori e assomiglia a una membrana «non otturabile» per ultrafiltrazione, se si eccettua la maggior densità dei gruppi carichi. La membrana solfonata consente il passaggio di ioni positivi, come quelli di sodio (Na^+), mentre blocca quelli negativi, come gli ioni cloro (Cl^-). La membrana carica positivamente *(a destra)* possiede selettività opposta. Le cariche positive sono fornite da un derivato dello ione ammonio quaternario (NH_4^+).

gior parte degli altri sali di piombo e bario sono quasi completamente insolubili.

La scelta dei polimeri solfonati come materiali costitutivi di membrane non suscettibili di otturarsi trova un'incoraggiante conferma in un esempio naturale: la sostanza colloidale nota come eparina secreta dalle tuniche dei vasi sanguigni. Si ritiene che la funzione dell'eparina, che ha la struttura di un polimero solfonato, sia quella di mantenere le pareti dei vasi sanguigni umide e libere da coaguli.

Sono state messe a punto tecniche per ottenere membrane di polimeri stirenici contenenti gruppi solfonici disposti sulla superficie della stessa membrana e all'interno dei pori. In genere, queste tecniche presuppongono la formazione di un sottile film di polimero solfonato e la successiva aggiunta di un agente chimico che favorisca la formazione di intrecci tra le lunghe catene delle molecole.

Si sono ottenuti risultati notevoli sperimentando membrane solfonate con soluzioni che generalmente producono formazione di massicci depositi otturanti. La procedura standard di prova per membrane, fino ad allora seguita, era stata quella di montare una membrana nell'apparecchiatura di prova e quindi di sottoporla all'azione di una singola soluzione fino a che la sua capacità filtrante non fosse gravemente compromessa. A questo punto, si procedeva all'installazione di una nuova membrana per la successiva soluzione di prova. Con le membrane solfonate fu invece possibile sottoporre un singolo campione di membrana all'azione successiva di molte soluzioni, note per la loro capacità di produrre depositi otturanti, per periodi di parecchi mesi, con trascurabili perdite nell'efficienza filtrante. Per esempio, una membrana fu tenuta sotto prova per nove mesi con una serie di liquidi comprendenti soluzioni di sali, coloranti, effluenti di industrie cartarie e di fabbriche di pasta di legno, melasse, siero proveniente dalla lavorazione del formaggio e acque luride. Il flusso attraverso la membrana variava per ogni soluzione, ma tale variazione era probabilmente provocata non tanto da fenomeni di otturazione, quanto da differenze nella retropressione osmotica. Al termine della serie di prove, la membrana risultava ancora pulita e brillante.

La resistenza all'otturazione manifestata dai polimeri solfonati ne allarga il campo di utilità oltre a quello della tecnologia basata su membrane. Uno di noi (Harry P. Gregor) ha preparato schermi e tubi rivestiti con questi polimeri e ha mostrato che essi rimangono relativamente puliti anche in condizioni sfavorevoli. Questo tipo di rivestimento suscettibile di non rivestirsi di depositi potrebbe impedire la fissazione di microrganismi marini su apparecchiature immerse per lungo tempo nel mare. Una possibile applicazione di tali rivestimenti potrebbe verificarsi nei dispositivi che fossero impiegati per sfruttare l'energia dei gradienti di temperatura nel mare. L'efficienza delle superfici di scambio termico facenti parte di questi dispositivi viene infatti sostanzialmente ridotta dall'eventuale formazione di depositi e incrostazioni.

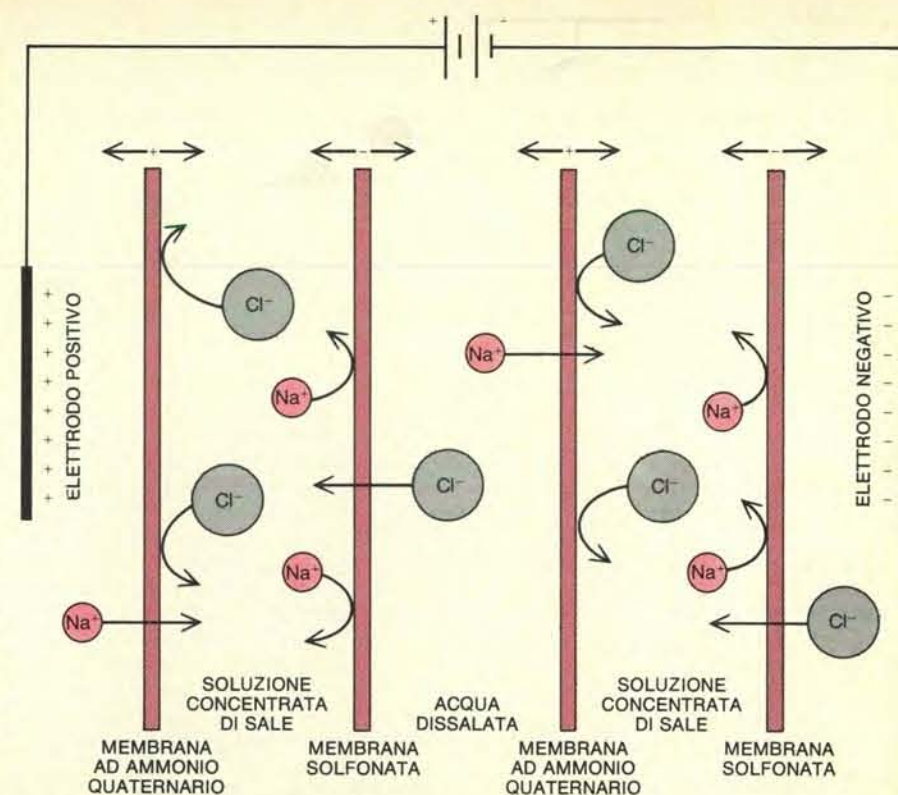
Oltre ad assicurare resistenza alla formazione di depositi otturanti, le cariche localizzate sui gruppi solfonici conferiscono un'altra prerogativa alle membrane, esaltandone la capacità di respingere sali. Infatti, le membrane elettricamente neutre, anche quelle con pori molto piccoli, hanno soltanto una limitata capacità di escludere i sali, presenti in soluzione come ioni positivi e negativi. Nelle membrane solfonate, invece, le cariche fisse negative producono un campo elettrico che respinge gli ioni con la carica dello stesso segno, impedendone l'accesso attraverso i pori della membrana. Poiché deve essere rispettata la regola per cui la soluzione deve mantenersi elettricamente neutra, vengono egualmente bloccati gli ioni positivi.

La capacità delle membrane cariche per ultrafiltrazione di respingere sali si conserva fino a quando la soluzione salina rimane relativamente diluita, poiché in una soluzione concentrata è presente una quantità di ioni positivi sufficiente a circondare e schermare la carica negativa dei gruppi solfonici, cosa che permette il passaggio agli ioni di entrambi i segni. L'esistenza di questo effetto di schermo rende impraticabile la tecnologia basata sull'ultrafiltrazione ai fini della dissalazione dell'acqua marina, che contiene circa 35 grammi di sali per litro. Tuttavia, nel caso delle acque salmastre, che contengono circa 5 grammi di sali per litro, è possibile arrivare a una riduzione dell'80 per cento della salinità. Un fatto più degno di interesse concernente le membrane cariche è la loro tendenza a respingere preferenzialmente gli ioni metallici bivalenti, come lo ione calcio Ca^{++} e lo ione magnesio Mg^{++}. Poichè questi ioni sono quelli che maggiormente contribuiscono alla durezza delle acque, le membrane possono venire utilmente impiegate nell'addolcimento di queste ultime.

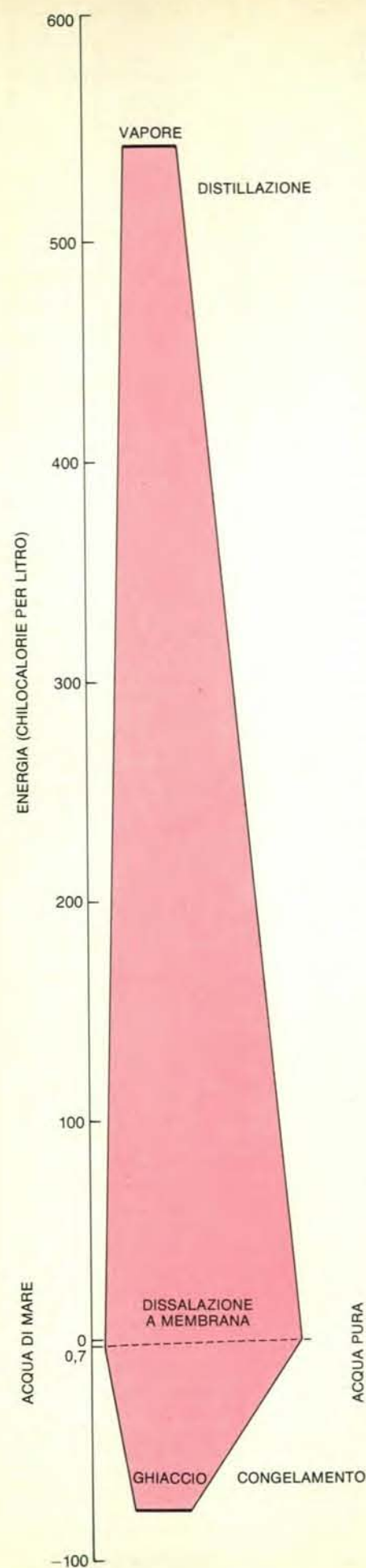
Rendendo sempre più piccoli i pori di una membrana polimerica si determina, alla fine, una condizione per cui la membrana è permeabile all'acqua e impermeabile a quasi tutto il resto. È possibile quindi estrarre acqua da una soluzione forzandola ad attraversare, sotto pressione, una simile membrana: questo processo è chiamato osmosi inversa. Come modalità di procedura, l'osmosi inversa potrebbe sembrare simile a un'ultrafiltrazione su scala estremamente ridotta, ma in realtà essa è presieduta da un meccanismo completamente diverso. In effetti in questo processo anche molecole piccole come quelle dell'acqua o di dimensioni inferiori a esse non riescono ad attraversare la membrana la quale non possiede pori definibili come tali. Anche il flusso di attraversamento dell'acqua non si compie nel senso consueto; al contrario, l'acqua è presente nella struttura della membrana in uno stato che viene definito simile al ghiaccio.

Nell'acqua allo stato liquido, a ogni istante, circa metà delle molecole sono addensate in raggruppamenti aventi il medesimo ordinamento strutturale che esiste in un cristallo di ghiaccio, come fu dimostrato per la prima volta nel 1933 da J. D. Bernal e R. M. Fowler. Nei raggruppamenti, ciascuna molecola d'acqua è disposta in modo che l'atomo di ossigeno occupi il vertice di un tetraedro e che un legame a idrogeno unisca ciascuna coppia di molecole. Nel ghiaccio, questa struttura stabile si estende su grandi distanze, mentre nello stato liquido ciascun raggruppamento simile al ghiaccio coinvolge generalmente solo poche molecole e inoltre viene a trovarsi in una continua condizione di formazione e disintegrazione.

In una membrana per osmosi inversa, un piccolo volume di acqua viene trattenuto negli interstizi tra le fibre costituenti la membrana dove assume una configurazione simile al ghiaccio. Applicando una pressione attraverso la membrana, le molecole che si trovano in corrispondenza del lato sotto pressione vengono incorporate nella struttura simile al ghiaccio, prendendo il posto di altre molecole che fuoriescono «per fusione» dal lato opposto. Le sostanze disciolte, con pochissime eccezioni, non sono in grado di adattarsi alla struttura simile al ghiaccio e quindi vengono estromesse. In particolare, gli ioni in soluzione sono circondati da molecole d'acqua disposte in modo da schermarne la carica, circostanza che impedisce agli ioni stessi di trovare sistemazione nella matrice simile al ghiaccio. Tra le poche eccezioni, si annoverano molecole capaci di instaurare legami a idrogeno con l'acqua, come il metanolo e l'urea; è significativo che queste sostanze siano in grado di penetrare, non solo attraverso le membrane per osmosi inversa, ma anche attraverso il comune ghiaccio.

Il fenomeno dell'osmosi inversa, pur essendo conosciuto da oltre un secolo, non trovò applicazioni pratiche a causa della permeabilità estremamente ridotta di queste membrane rigide che consentivano un flusso di acqua estremamente scarso. La soluzione a questo problema fu quella di realizzare membrane più sottili: come è stato sottolineato in precedenza, il flusso di solvente è inversamente proporzionale allo spessore della membrana, mentre lo spessore non ha influenza sulla facoltà di respingere i soluti. Non appariva comunque attuabile la possibilità di produrre membrane sufficientemente sottili da consentire valori di flusso di utilità pratica.

In elettrodialisi si impiegano membrane a cariche fisse per estrarre acqua pura da soluzioni saline. A differenza dell'ultrafiltrazione e dell'osmosi inversa, l'elettrodialisi non è promossa da una pressione applicata, bensì da un campo elettrico. Le membrane sono cotraddistinte *(in alto)* in base alla carica ionica che fanno passare. Gli ioni cloro sono attratti dall'elettrodo positivo e pertanto si muovono verso sinistra, passando attraverso ogni membrana solfonata che incontrano sul proprio cammino, ma venendo bloccati da quelle contenenti derivati dall'ammonio quaternario. Analogamente, gli ioni sodio si muovono verso destra finché non vengono arrestati da una membrana solfonata. Come risultato, si ha un accumulo di acqua pura nella cavità centrale, mentre la soluzione salina si concentra negli spazi tra le due coppie di membrane periferiche.

Nel 1960, Sidney Loeb dell'Università della California a Los Angeles (ora presso l'Università Ben Gurion in Israele) scoprì un sistema per produrre membrane che fossero al contempo sufficientemente sottili e rigide da essere utilizzabili nella pratica dell'osmosi inversa. Il suo sistema sfrutta un fenomeno che si osserva quando si ottiene una membrana lasciandone evaporare il solvente: dopo un periodo iniziale di essiccamento, si forma una «buccia» dovuta al fatto che la soluzione presente nella parte superficiale evapora più rapidamente di quella presente negli strati sottostanti. In genere, si cerca di evitare la formazione di questa buccia per ottenere una membrana di porosità uniforme. Loeb, invece, lasciò deliberatamente che si formasse la buccia e quindi immerse la membrana in acqua per fare rapprendere il polimero degli strati più profondi che non si era ancora completamente essiccato. Il risultato fu una membrana costituita da una pellicola estremamente sottile di polimero rigido e coerente, sovrapposta a uno strato assai più spesso di materiale di elevata porosità. Membrane di questo tipo sono sufficientemente robuste per resistere a una pressione elevata, mentre la loro buccia attiva ha uno spessore che corrisponde all'uno per cento circa di quello di una comune membrana per ultrafiltrazione. Conseguentemente, il flusso di acqua è 100 volte superiore.

Una delle principali applicazioni dell'osmosi inversa si ha nell'estrazione di acqua pura da soluzioni saline, in particolare da acque salmastre e acque di mare. Diversamente dalle membrane per ultrafiltrazione, le membrane per osmosi inversa non diventano permeabili agli ioni a elevate concentrazioni di sali, ma esiste tuttavia un limite nel grado di concentrazione a cui può venire portata la soluzione di alimentazione. Per avere flussi di acqua di interesse pratico, la pressione netta attraverso la membrana deve essere di almeno alcune decine di chilogrammi per centimetro quadrato. Quando la concentrazione della soluzione salina raggiunge il valore del 3 per cento circa, la pressione osmotica arriva a toccare i 28 chilogrammi per centimetro quadrato. Il costo per il funzionamento di pompe e altri dispositivi vincola la pressione applicata a un limite di circa 70 chilogrammi per centimetro quadrato, oltre al quale ogni tentativo di aumentare la concentrazione diventa antieconomico. È possibile estrarre acqua pura dal mare per osmosi inversa se i sali trattenuti possono essere illimitatamente diluiti in un serbatoio assai esteso (l'oceano per l'appunto). Se invece è necessario

concentrare i sali, sia per utilizzarli che per eliminarli, l'osmosi inversa deve essere abbinata ad altri processi che non siano limitati nella propria efficienza dalla concentrazione della soluzione che alimenta il sistema.

La dissalazione per osmosi inversa mette a fuoco una proprietà comune a molti processi basati su membrane: il fatto di richiedere meno energia di altri sistemi di separazione confrontabili, basati sul calore. L'estrazione di un litro di acqua pura da un volume illimitato di acqua di mare richiede un minimo di 0,7 chilocalorie, la minima energia teoricamente necessaria da un punto di vista termodinamico, che costituisce un valore limite indipendentemente dal sistema di separazione impiegato. Il minimo termodinamico può essere avvicinato solo quando tutti gli stadi costituenti un processo sono reversibili, cioè quando si svolgono in condizioni vicine all'equilibrio. È possibile realizzare l'evaporazione e la condensazione in una situazione assai vicina alla reversibilità, facendo avvenire entrambi i processi alla stessa temperatura e pressione. In queste condizioni di equilibrio, tuttavia, i processi termici sono eccessivamente lenti e richiedono apparecchiature di enormi dimensioni. Se ne può accrescere la velocità operando con una rilevante differenza di temperatura, ma questo scostamento dall'equilibrio e dalla reversibilità impone un pesante costo energetico. L'estrazione di un litro di acqua per distillazione richiede, per esempio, qualcosa come 540 chilocalorie, di cui solo una parte può essere recuperata in uno scambiatore di calore. Il funzionamento delle membrane, d'altra parte, si compie in condizioni apprezzabilmente vicine all'equilibrio, come sta a dimostrare il fatto che lo stesso litro di acqua può essere separato per osmosi inversa con un investimento di circa tre chilocalorie, soltanto quattro volte o poco più rispetto al minimo di energia.

Ciò che determina sia l'ultrafiltrazione che l'osmosi inversa è la pressione applicata, ma l'energia necessaria per la separazione di molecole mediante membrana può venire anche da altre fonti. Nel processo chiamato elettrodialisi l'energia viene fornita da un campo elettrico applicato attraverso una membrana o una serie di membrane. Nella dialisi, è il soluto che si muove attraverso la membrana, anzichè il solvente; nell'elettrodialisi, i soluti sono ioni e le membrane sono selettivamente permeabili agli ioni carichi, sia positivi che negativi.

Nel 1923, il biochimico tedesco Leonor Michaelis dimostrò che le membrane prodotte con nitrocellulosa essiccata sono selettivamente permeabili agli ioni positivi, ma non seppe spiegare il meccanismo di questo fenomeno. Nel 1929, Karl Sollner iniziò il primo studio sistematico in cui veniva messa in relazione la struttura delle membrane con il loro funzionamento. Insieme a due suoi studenti, Charles Carr e uno di noi (Harry P. Gregor), Sollner fabbricò le prime membrane progettate in modo che potessero operare non solo con elevata selettività, ma anche con permeabilità ottimale.

Attualmente, la tecnologia dell'elettrodialisi è ben sviluppata e le relative membrane vengono prodotte in quantità commerciali. Le membrane possiedono pori estremamente fini (del diametro di circa un nanometro) e vengono ottenute da polimeri reticolati, sui quali si localizzata una elevata concentrazione di cariche fisse. Le membrane cariche negativamente sono fabbricate con polistirolo a cui sono legati gruppi solfonici e questo consente loro di respingere tutti gli ioni con carica negativa. Pur essendo simili alle membrane solfonate per ultrafiltrazione, hanno una permeabilità all'acqua molto inferiore e una maggiore concentrazione di cariche fisse. Grazie all'elevata densità di gruppi carichi elettricamente, queste membrane sono altamente selettive anche in presenza di soluzioni saline con concentrazioni fino a 175 grammi per litro.

Le corrispondenti membrane cariche positivamente sono fabbricate con il polimero polivinil-benzilcloruro e la carica necessaria è fornita da gruppi derivati dallo ione ammonio quaternario (NH_4^+), come per esempio il trimetilammonio. Com'è da attendersi, la selettività di queste membrane è rivolta agli ioni positivi, mentre gli ioni negativi hanno libertà di passaggio. Le membrane solfonate, al pari delle analoghe impiegate in ultrafiltrazione, mostrano un'elevata resistenza all'otturazione, cosa che invece non si verifica per quelle contenenti ammonio quaternario. Queste ultime, infatti, si otturano facilmente se vengono esposte a fluidi contenenti prodotti naturali e questo inconveniente rende talvolta necessario, prima dell'elettrodialisi, un pretrattamento della soluzione ottenuto per ultrafiltrazione.

In una cella di elettrodialisi, membrane di polarità alternata sono sistemate in serie e ai capi dell'insieme viene applicato un campo elettrico. Gli ioni positivi sono attratti dall'elettrodo negativo su cui si dirigono passando attraverso una membrana solfonata, ma sono comunque arrestati dalla successiva membrana, del tipo ad ammonio quaternario. Gli ioni negativi si muovono in direzione opposta, fino a che non vengono fermati da una membrana solfonata. Come risultato, gli scomparti tra le membrane si riempiono alternativamente di acqua pura e di soluzione salina concentrata.

L'energia necessaria per la dissalazione dell'acqua di mare è teoricamente indipendente dalle modalità di estrazione, ma, nella pratica, differenze nell'efficienza del processo inducono sensibili differenze nel fabbisogno energetico. La quantità minima teorica di energia che occorre spendere per ottenere un litro di acqua pura da un grande volume di acqua di mare è di 0,7 chilocalorie. Nella distillazione è necessario innalzare inizialmente l'energia al livello di 540 chilocalorie, ma, con un perfetto scambiatore di calore, l'intera quantità, eccetto 0,7 chilocalorie, potrebbe venire recuperata. Nella realtà, ne va perduta una parte rilevante. Similmente, la separazione di sale per congelamento comporta un investimento di —79 chilocalorie, solo in piccola parte recuperabili. Poiché la dissalazione mediante membrana non è un processo termico, non è necessario un tale investimento energetico e l'efficienza è assai elevata. Il consumo è di circa 3 chilocalorie per litro, quattro volte il minimo teorico.

In Giappone, la maggior parte del sale da cucina viene estratto dall'acqua di mare con un metodo che comprende uno stadio di elettrodialisi. Dopo che gli ioni magnesio e solfato sono stati allontanati con un altro processo che utilizza membrane, la soluzione di cloruro di sodio viene concentrata circa sei volte rispetto a quella di partenza mediante elettrodialisi. L'acqua residua viene eliminata per evaporazione. L'energia necessaria per l'elettrodialisi ammonta a circa 300 chilowattora per tonnellata di sale, mentre altri 250 chilowattora vengono consumati nell'evaporazione dell'acqua. Se l'intero processo di concentrazione fosse compiuto per evaporazione, occorrerebbero 2100 chilowattora per tonnellata, il che dimostra che il ricorso alla tecnica ibrida permette di consumare soltanto un quarto circa dell'energia.

Nel 1913, Albert Bethe (fisiologo e fratello del fisico Hans Bethe) applicò una membrana animale tra due soluzioni e fece passare una corrente elettrica attraverso il sistema, allo scopo di osservare il movimento delle molecole di acqua. Con sorpresa scoprì che la soluzione, da un lato della membrana, diveniva debolmente acida, mentre dall'altro lato diveniva debolmente alcalina. Questo effetto fu da lui attribuito alla dissociazione delle molecole d'acqua in ioni idrogeno (H^+) e ossidrile (OH^-). Il fenomeno scoperto da Bethe rimase una curiosità di laboratorio fino al momento in cui si constatò che anche le membrane commerciali per elettrodialisi determinavano la dissociazione di apprezzabili quantità di acqua. Emerse in seguito con evidenza la possibilità di impiegare le membrane per ottenere un acido e una base a partire da un sale.

È importante sottolineare la distinzione tra questo tipo di dissociazione dell'acqua e l'apparentemente simile processo dell'elettrodialisi, in cui i prodotti finali sono idrogeneo (H_2) e ossigeno (O_2) molecolare. La dissociazione della molecola di acqua negli atomi che la compongono richiede quantità di energia assai elevate e non riducibili, mentre la dissociazione ionica della stessa molecola avviene spontaneamente, senza alcun contributo di energia. Nell'acqua pura, a ogni istante, circa una molecola ogni 10^7 è ionizzata. Nella tecnica di dissociazione dell'acqua mediante membrana la corrente elettrica serve essenzialmente a tenere separati gli ioni dopo che essi si sono formati, anche se, come si è recentemente scoperto, la presenza del campo elettrico accentua debolmente la ionizzazione.

Una cella per la dissociazione dell'acqua può essere realizzata con membrane del tutto simili a quelle utilizzate in elettrodialisi. Due membrane di opposta polarità sono saldate l'una contro l'altra e sono poi circondate da altre due coppie di membrane, ma per ciascun lato, sempre con polarità alternata. L'intera cella è immersa in una soluzione di un sale, quale cloruro di sodio, e il sistema viene attraversato da una corrente elettrica. Gli ioni idrogeno e ossidrile che si formano nell'interstizio tra le membrane saldate sono separati in base alla carica; essi attraversano una membrana e sono respinti dalla successiva, per cui si combinano con gli ioni sodio e cloro intrappolati tra le membrane. Da un lato della doppia membrana si forma acido cloridrico (HCl), dall'altro lato idrossido di sodio (NaOH).

Nelle prime membrane per la dissociazione dell'acqua si manifestava un problema legato alla disseccazione della doppia membrana quando gli stessi ioni idrogeno e ossidrile, insieme alle loro molecole d'acqua di idratazione, venivano estratti dall'interstizio. Date le elevate densità di corrente richieste per le applicazioni industriali le membrane diventavano presto inutilizzabili. Recentemente, uno di noi (Harry P. Gregor), insieme con Kenneth Brennen e Bruce Benjamin, ha messo a punto metodi per impedire la disseccazione attraverso l'accurata giunzione di membrane selezionate.

La tecnologia basata sulla dissociazione dell'acqua, pur essendo ancora in corso di perfezionamento, possiede già numerose applicazioni promettenti. Nell'estrazione mineraria di metalli, per esempio, si potrebbero produrre direttamente in loco l'acido e la base: il primo potrebbe venire impiegato nell'estrazione del metallo dal minerale, mentre la base potrebbe essere destinata alla neutralizzazione dei liquidi residui prima del loro scarico.

In una modificazione della tecnica basata sulla decomposizione dell'acqua, viene dissociato un sale di un acido e una base deboli nei suoi componenti gassosi. Il sale di solfuro di ammonio, per esempio, è un sottoprodotto della raffinazione del petrolio, che si forma nel trattamento di desolforazione del greggio con ammoniaca gassosa. Una membrana del tipo impiegato nella dissociazione dell'acqua potrebbe decomporre questo prodotto di rifiuto, permettendo il recupero di ammoniaca e solfuro di idrogeno, che sono prodotti di un certo pregio.

Un terzo criterio per operare separazioni tra molecole, oltre alle loro dimensioni e alla loro carica elettrica, è la loro solubilità in solventi differenti. Le differenze nella solubilità sono sfruttate nel processo detto di estrazione con solvente. Nella sua forma più semplice, l'estrazione con solvente consiste nel purificare una sostanza disciolta in acqua agitando la

Nelle celle per la dissociazione dell'acqua, si utilizzano membrane a cariche fisse per ottenere un acido e una base a partire da una soluzione salina. Le due membrane al centro sono saldate insieme, ma tra l'una e l'altra esiste un ridotto spazio che può accogliere acqua. Come sempre, nell'acqua una piccola frazione di molecole (circa 10^{-7}) si ionizza spontaneamente. Gli ioni vengono quindi separati dal campo elettrico, in modo che gli ioni idrogeno (H^+) vanno verso destra e gli ioni ossidrile (OH^-) verso sinistra. Nel contempo, si concentrano ioni sodio (Na^+) nel comparto a sinistra della doppia membrana e ioni cloro (Cl^-) a destra. La ricombinazione tra ioni provenienti dall'acqua e ioni provenienti dal sale dà idrossido di sodio e acido cloridrico.

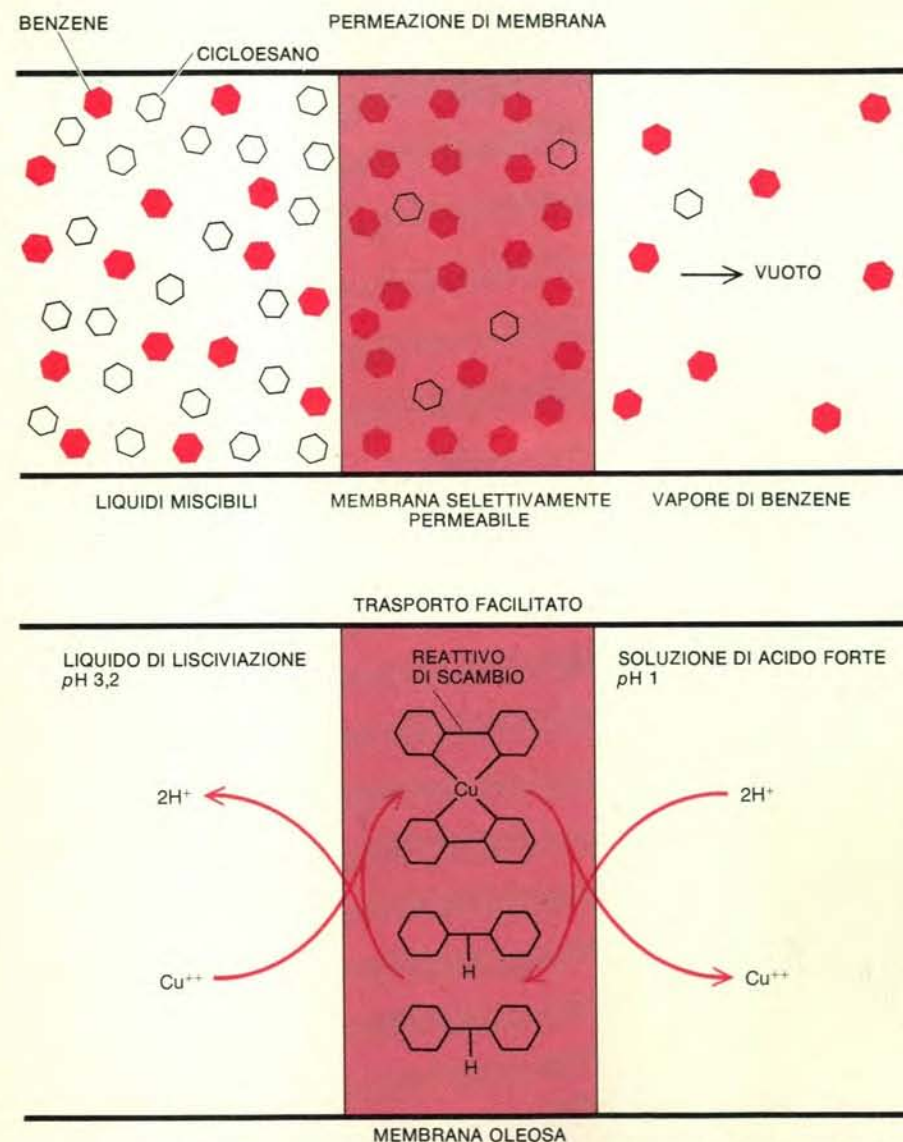
soluzione con un altro solvente, per esempio olio, non miscibile con acqua, ma nel quale la sostanza sia più solubile.

L'estrazione con solvente attraverso membrane era nota almeno fin dal 1913, quando Fritz Haber e Reinhardt Beutner dimostrarono che una sottile pellicola di olio poteva essere impiegata come una membrana in due tipi di processi di estrazione. Nel primo, chiamato permeazione attraverso membrana, una miscela di due liquidi miscibili, quali benzene e cicloesano, viene posta da un lato della membrana, mentre dall'altro lato viene praticato il vuoto. Se la membrana è più permeabile a un componente, per esempio il benzene, il vapore che fuoriesce dalla parte del vuoto è prevalentemente benzene. Diversamente dalla distillazione, la permeazione attraverso membrana richiede l'evaporazione di un solo componente di una miscela, riducendo quindi il fabbisogno di energia.

Un'altra forma di estrazione con solvente ideata da Haber e Beutner, chiamata trasporto facilitato, assomiglia, per certi versi, al meccanismo che si presume presieda al trasporto attivo di molecole attraverso membrane biologiche. Da entrambi i lati di una membrana oleosa vengono poste soluzioni acquose, una delle quali contiene la sostanza che si vuole isolare. Una molecola mobile della membrana forma quindi un complesso con la sostanza da isolare e il complesso così ottenuto diffonde attraverso la membrana e si dissocia dalla parte opposta.

Lawrence Brandlein e uno di noi (Harry P. Gregor) hanno dimostrato che, con questo metodo, è possibile estrarre il rame dalle soluzioni in cui viene raccolto per lisciviazione dal minerale. Il rame è presente, con altri metalli, in una soluzione debolmente acida che viene posta da un lato della membrana impregnata con una soluzione di cherosene. Gli ioni metallici non sono solubili in cherosene, per cui la membrana è a essi impermeabile. Gli ioni di rame sono incorporati nella membrana da un reagente organico che forma selettivamente con il rame (escludendo gli altri metalli) un complesso stabile solubile in cherosene. Dall'altra parte della membrana, il rame viene spostato dal complesso per opera di un acido forte. In questo modo, il rame, in una singola operazione, viene estratto dal liquido di lisciviazione e protetto dalla contaminazione di altri metalli. L'estrazione con solvente attraverso membrane potrebbe anche venire applicata al ritrattamento dei combustibili nucleari.

Rimane un ulteriore processo basato su membrane, dove non vengono però sfruttate le loro facoltà selettive, in quanto le membrane fungono sostanzialmente da matrice inerte su cui vengono fissati enzimi. Nel 1954, Nikolaus Grubhofer dimostrò che era possibile accoppiare chimicamente un enzima a una matrice sintetica senza che esso divenisse inattivo. Un successivo lavoro di Ephraim Katchalsky e Georg Manecke consentì la realizzazione di reattori chimici dove gli enzimi sono chimicamente legati a letti porosi e insolubili. (Katchalsky, che più tardi prese il nome di Katzir, si è da poco ritirato dall'incarico di presidente di Israele). Questo tipo di reattore trova ora impiego commerciale (per esempio nella trasformazione di glucosio in zucchero invertito), ma presenta taluni difetti. Per promuovere il legame tra l'enzima e il letto poroso è generalmente necessaria la presenza contemporanea dell'enzima e di un agente copulante corrosivo, col risultato che una buona parte dell'enzima viene spesso distrutta. Inoltre, poiché i reagenti e i prodotti circolano solo per diffusione, il processo è piuttosto lento.

Paul Rauf e uno di noi (Harry P. Gregor) hanno dimostrato che queste difficoltà possono essere superate attraverso l'accoppiamento dell'enzima con membrane aventi pori di dimensioni circa 5 volte maggiori delle molecole enzimatiche. La preparazione della membrana viene effettuata pompando attraverso i pori un agente copulante, quale bromuro di cianogeno (CNBr), che forma un intermedio instabile capace di legare l'enzima. Si procede quindi all'eliminazione,

L'estrazione con solvente attraverso una membrana permette di separare molecole in base alla loro diversa affinità per un solvente o per una molecola specializzata nel trasporto, presente nella membrana. In un tipo di estrazione con solvente, chiamato permeazione attraverso membrana *(in alto)*, una miscela di due liquidi miscibili, per esempio benzene e cicloesano, viene collocata da un lato della membrana che assorbe più agevolmente un componente della miscela rispetto all'altro. Praticando quindi un vuoto sul lato opposto della membrana, vengono trascinate attraverso quest'ultima soprattutto molecole del componente più permeante, nel nostro caso benzene. Un secondo procedimento di estrazione, detto trasporto facilitato *(in basso)*, può essere sfruttato per estrarre rame dalla soluzione acida con cui il metallo viene lisciviato dal minerale. La membrana usata, pur essendo impermeabile al rame (come ad altri ioni metallici), incorpora tuttavia un reattivo di scambio in grado di formare selettivamente un complesso con il rame. Il complesso si forma sulla superficie della membrana prospiciente la soluzione concentrata e diffonde verso la parte opposta, dove il rame viene spostato da un acido forte, mentre il reattivo si rigenera.

per lavaggio, dell'agente copulante, dopo di che viene introdotto l'enzima sotto pressione. A condizione di controllare la pressione e il tempo di reazione, è possibile ricoprire i pori di un denso strato di enzima, che conserva fino all'80 per cento della sua attività originaria. La fissazione dell'enzima alla membrana è essenzialmente una «saldatura a punti» e l'attività enzimatica delle molecole subisce interferenze soltanto nel caso in cui la saldatura dovesse instaurarsi in una posizione vicina al sito attivo.

In un esperimento, si è impregnata una membrana con l'enzima chimotripsina che, oltre a essere normalmente capace di digerire proteine, è anche in grado di decomporre un estere in un acido e in un alcool. Si è trovato che la velocità di decomposizione degli esteri aumenta con la pressione e che l'attività catalitica dell'enzima raggiunge la sua piena capacità solo quando la pressione applicata tocca valori di circa 7 chilogrammi per centimetro quadrato.

Sono state qui descritte numerose applicazioni tecnologiche di membrane sintetiche, ma esistono altre applicazioni potenziali, tra le quali vale la pena di illustrare, a titolo esemplificativo, quelle che si riferiscono a tre generi di attività industriali suscettibili di adottare membrane su scala particolarmente estesa. Ci riferiamo ai trattamenti degli effluenti di fogna, all'industria alimentare e all'eventuale coltivazione di piante verdi come fonte di combustibile.

Il trattamento convenzionale delle acque di fogna prevede l'impiego combinato di serbatoi di decantazione, colture batteriche e sistemi di ispessimento dei fanghi allo scopo di decontaminare le acque di scarico e di concentrare i residui solidi. Il trattamento preliminare, che permette la decantazione delle particelle più voluminose, non può probabilmente trarre miglioramento da tecnologie basate su membrane. Il secondo tipo di trattamento, quello con batteri, potrebbe essere sostituito dall'ultrafiltrazione con membrane non suscettibili di otturarsi. Come hanno mostrato Kang Hsu, Sadi Mizrahi e uno di noi (Harry P. Gregor), gli effluenti provenienti da tale tipo di filtrazione sono di qualità uguale o superiore a quella ottenibile nel trattamento convenzionale. Infatti, poiché batteri e virus non possono penetrare attraverso la membrana, la soluzione filtrata è pressoché sterile. L'aspetto più importante è che i solidi trattenuti non vengono decomposti, per cui essi decantano più rapidamente e possono venire più facilmente concentrati agevolandone l'eliminazione. I fanghi concentrati sono in grado di mantenere la propria combustione in un inceneritore e, inoltre, essendo costituiti per circa il 20-30 per cento da proteine, rappresentano una fonte potenziale di fertilizzanti azotati.

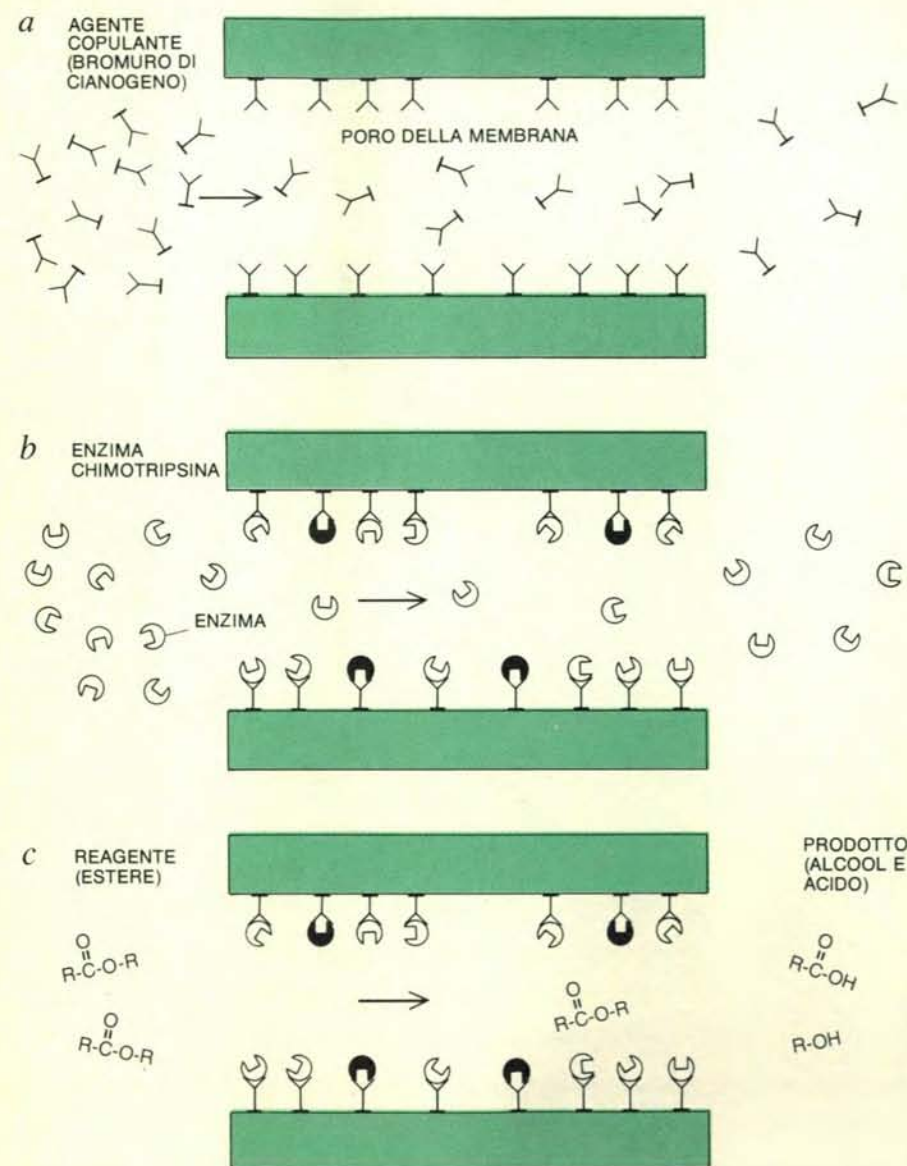
Poiché le unità di ultrafiltrazione sono modulari, sarebbero economicamente realizzabili piccoli impianti di trattamento. Viene infatti soppesata la possibilità di impiegare l'ultrafiltrazione in appoggio al trattamento con fosse settiche per edifici singoli. L'ultrafiltrazione potrebbe rivelarsi particolarmente vantaggiosa nel trattamento delle acque di rifiuto in regioni desertiche, dove, insieme ad altri contaminanti, occorre eliminare i sali.

Nella preparazione e nella raffinazione di molti alimenti, quali formaggio o soia, si formano rifiuti liquidi che contengono grandi quantità di sostanze nutritive, anche se in concentrazione troppo bassa per giustificarne il recupero in termini economici. D'altra parte, la semplice discarica di questi liquidi può provocare gravi problemi di inquinamento.

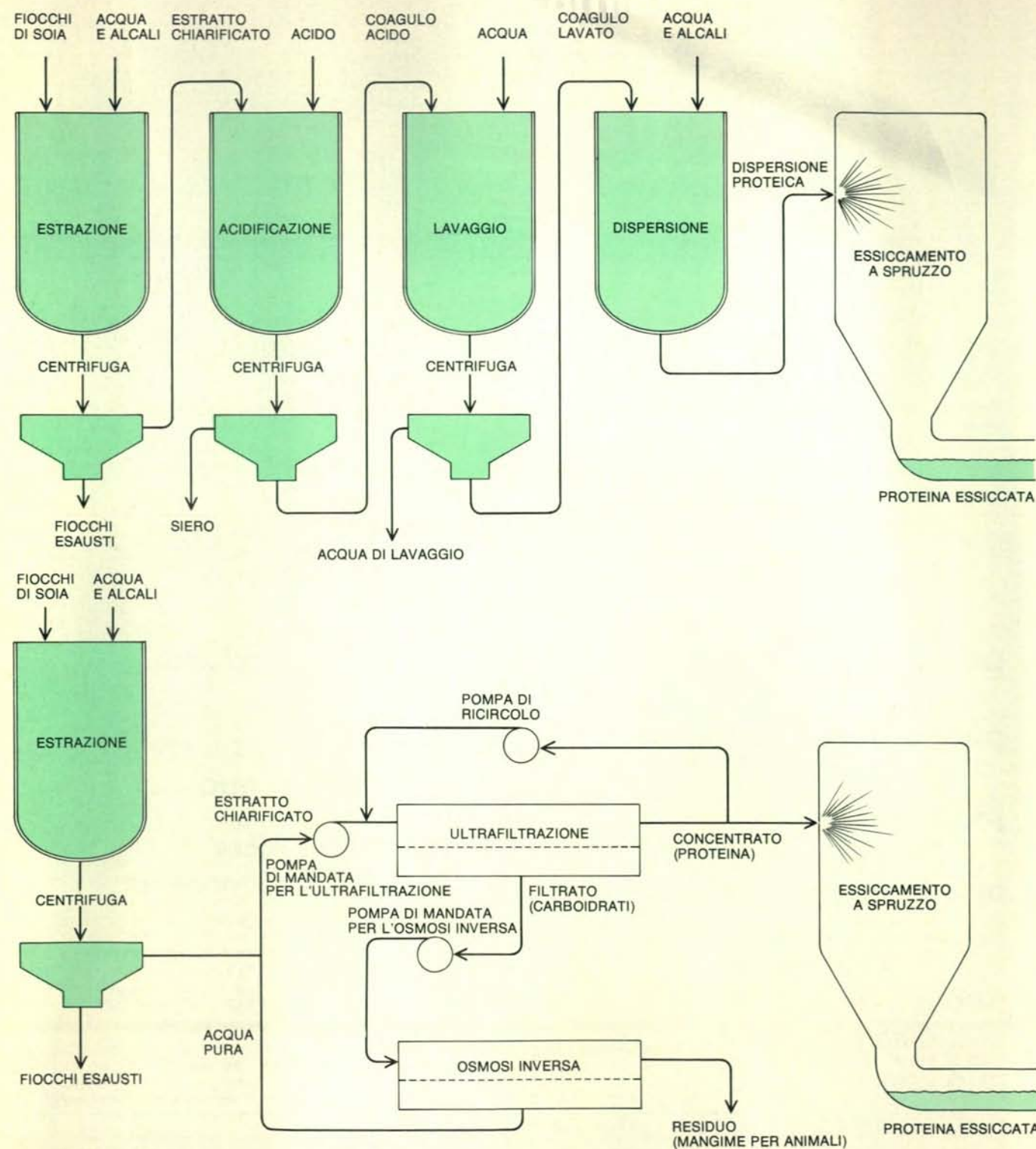
Nella fabbricazione del formaggio il liquido di rifiuto è il siero, che nei soli Stati Uniti è prodotto in quantitativi dell'ordine di 20 miliardi di litri all'anno. Il siero contiene circa il 3 per cento della proteina del latte lattalbumina e quantità inferiori di cenere e acido lattico. Può venir concentrato per evaporazione ed essiccato per essere usato come mangime animale, ma l'energia richiesta per l'intero processo lo rende economicamente poco vantaggioso. Recenti studi mostrano, tuttavia, che l'ultrafiltrazione del siero permette di ottenere un concentrato costituito dal 10-20 percento di proteine e un filtrato limpido contenente lattosio,

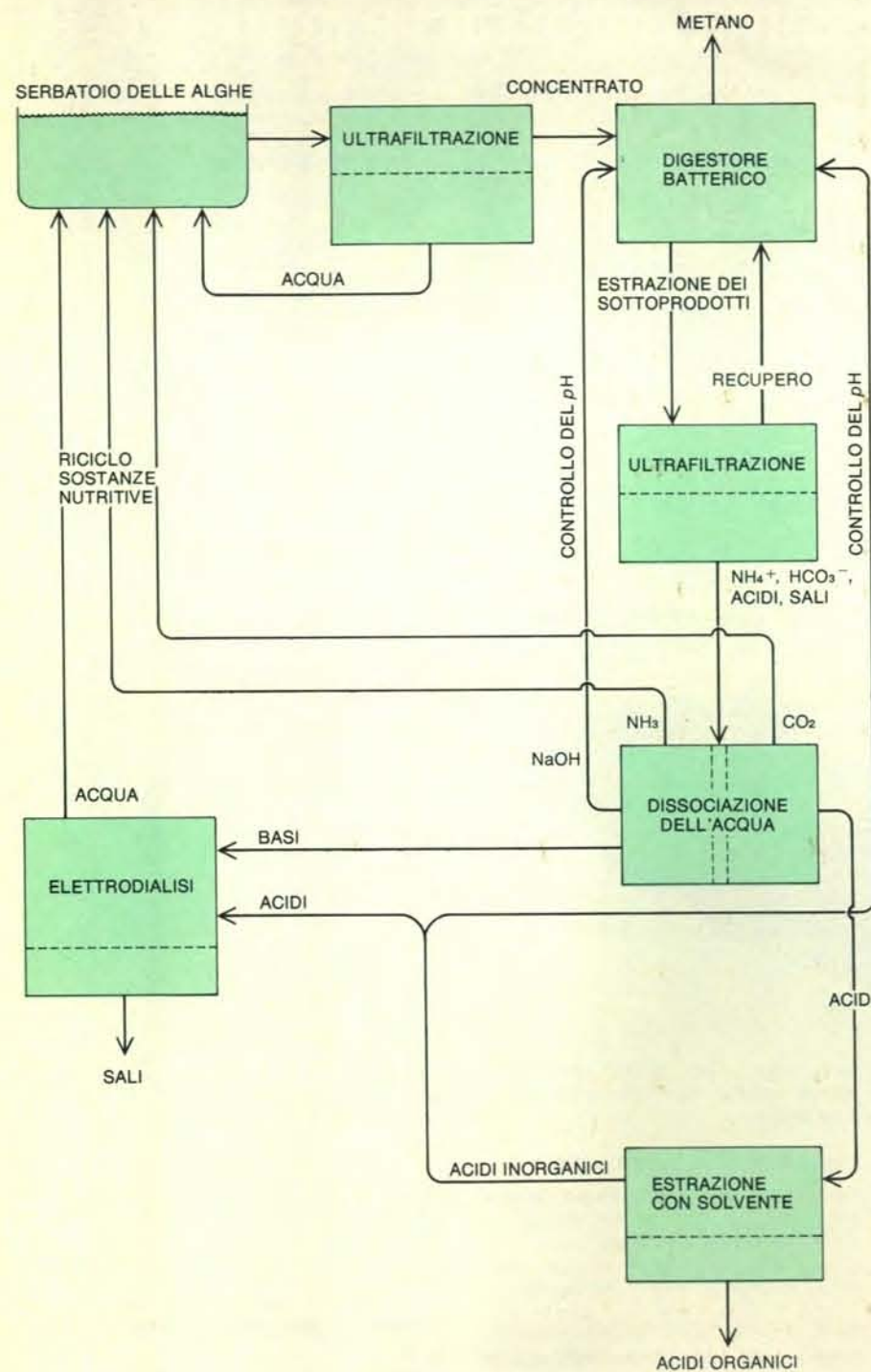
a AGENTE COPULANTE (BROMURO DI CIANOGENO) — PORO DELLA MEMBRANA

b ENZIMA CHIMOTRIPSINA — ENZIMA

c REAGENTE (ESTERE) — PRODOTTO (ALCOOL E ACIDO)

Un reattore enzimatico può essere costruito fissando chimicamente un enzima nei pori di una membrana polimerica. Dapprima, attraverso la membrana viene pompato un agente copulante, quale bromuro di cianogeno (CNBr), che si lega alle pareti dei pori *(a)*. Viene quindi introdotto, sempre sotto pressione, l'enzima, in questo caso l'enzima della digestione chimotripsina, che si lega all'agente copulante *(b)*. Alcune molecole di enzima vengono disattivate, principalmente quelle che casualmente si legano all'agente copulante in prossimità del proprio sito attivo, ma almeno l'80 per cento dell'attività enzimatica può essere conservata. Una volta che la membrana ha subito il trattamento preparatorio, si può pompare attraverso di essa un substrato con l'enzima, mentre i prodotti della reazione enzimatica fuoriescono dall'altro lato *(c)*. La chimotripsina, per esempio, è un enzima in grado di scindere un estere nell'acido organico e nell'alcool corrispondenti.

sali e acido lattico. Il lattosio può venire ulteriormente concentrato, per osmosi inversa, fino a circa il 24 per cento, per poi essere recuperato allo stato puro per evaporazione. In alternativa, la soluzione di lattosio può essere fatta passare attraverso un reattore enzimatico che lo trasforma negli zuccheri glucosio e galattosio, i quali risultano più gradevoli al gusto che non il lattosio.

La tecnologia basata su membrane può trovare analoghe applicazioni nella purificazione di altri residui e sottoprodotti dei processi di raffinazione alimentare. Per esempio, l'estrazione delle pro-

Il processo di estrazione di proteine dalla soia potrebbe essere semplificato sfruttando la tecnologia a membrane sintetiche. Nel comune metodo di estrazione *(in alto)*, le proteine vengono disciolte in un mezzo alcalino, insieme ad altre sostanze che comprendono alcuni carboidrati indesiderati. Le proteine vengono quindi coagulate per trattamento con acido, lavate in acqua per allontanare il grosso delle impurezze, disperse nuovamente in alcali e infine sottoposte a essiccamento a spruzzo. A ogni stadio, i residui in sospensione vengono rimossi per centrifugazione. Anche il processo mediante membrana inizia con un'estrazione con alcali, ma le sostanze estranee vengono separate per ultrafiltrazione. Il liquido filtrato può quindi essere sottoposto a osmosi inversa per concentrare i residui, costituiti essenzialmente da carboidrati, e per recuperare acqua pura. Il processo mediante membrana richiede meno energia e lascia una quantità più limitata di residui.

teine della farina di soia comporta la formazione di liquidi contenenti sia le proteine richieste che carboidrati e sali di rifiuto; la separazione e la purificazione di questi componenti potrebbero essere realizzate con una combinazione di ultrafiltrazione e osmosi inversa. La macinazione a umido del mais produce un liquido residuo chiamato, negli Stati Uniti, *high-solubles water;* un tipico impianto di

La coltivazione di alghe per l'ottenimento di combustibile, ancora allo stadio di proposta, è una tecnologia che potrebbe utilizzare le membrane sintetiche. Le alghe potrebbero essere raccolte per mezzo di una membrana per ultrafiltrazione con pori relativamente ampi e la sospensione concentrata di cellule vegetali così ottenuta potrebbe alimentare un digestore batterico, dove alla fine verrebbe prodotto il combustibile (per esempio metano). Un'altra membrana per ultrafiltrazione provvederebbe ad allontanare i sottoprodotti della digestione batterica e una cella per la dissociazione dell'acqua ne assicurerebbe la degradazione. L'ammoniaca (NH_3) e l'anidride carbonica (CO_2) rappresentano sostanze nutritive di pregio che potrebbero venire riciclate nello stagno o nel serbatoio di coltivazione delle alghe. La separazione di acidi organici, aventi valore commerciale, potrebbe essere effettuata per estrazione con solvente mediante membrana. Ricorrendo infine all'elettrodialisi, si potrebbero concentrare acidi inorganici e basi sotto forma di sali.

grandi dimensioni ne scarica quasi 80 milioni di litri al giorno, che vengono attualmente evaporati, a un costo di circa 200 000 dollari al giorno, fino a una concentrazione del 50 per cento ed essiccati a spruzzo direttamente sui foraggi insilati, per essere usati come mangime animale. Con l'ultrafiltrazione, si potrebbero allontanare tutto il materiale in sospensione e la sostanza proteica disciolta e recuperare il 99 per cento dell'acqua, che potrebbe venire successivamente purificata ad acqua potabile per osmosi inversa. Solo il residuo di questo secondo trattamento, contenente il 13 per cento di sali, richiede un'ulteriore concentrazione per evaporazione, al fine di essere impiegato come mangime animale. Da una valutazione economica del processo a membrana risulta che esso richiede solo il 20 per cento dell'energia necessaria per l'evaporazione diretta del liquido di rifiuto.

Un promettente sistema per catturare energia solare è legato alla coltivazione di specie vegetali fotosinteticamente efficienti che possono essere trasformate in un combustibile liquido, come per esempio metanolo (si veda l'articolo *La conversione biologica dell'energia solare* di I. F. Quercia e P. Quercia in «Le Scienze» n. 119, luglio 1978). Tra le piante che possono essere considerate adatte per ottenere combustibile, figurano le alghe, che sono efficienti trasformatrici di energia solare e che possono essere coltivate in numerosi ambienti. Esistono, tuttavia, due non trascurabili ostacoli a questo tipo di approvvigionamento di energia: da una parte, le cellule delle alghe, le cui dimensioni vanno da un micrometro a 10 micrometri, devono poter essere raccolte in modo rapido non dispendioso dagli stagni in cui venissero fatte crescere e, dall'altra, è necessario conservare i sali nutritivi, che vanno somministrati in quantitativi enormi per assicurare la crescita delle piante. La tecnologia basata su membrane potrebbe semplificare entrambi questi problemi.

Le cellule delle alghe sono presenti nell'acqua con una densità di solo 0,1-0,2 grammi per litro (peso a secco) e ciò ne rende necessaria la concentrazione a valori di almeno 10 grammi per litro prima che possano essere introdotte in un digestore batterico per essere tasformate in combustibile. Le alghe possono essere concentrate per coagulazione chimica o per centrifugazione, ma il costo di questi sistemi appare proibitivo se riferito a un tipo di raccolto che, essendo destinato essenzialmente a essere bruciato, dev'essere molto economico. Un'alternativa molto promettente è offerta dall'ultrafiltrazione, tanto più che, essendo le cellule piuttosto grandi in rapporto alla scala dei diametri dei pori, può essere impiegata una membrana «a maglie larghe», con elevata velocità di flusso. In esperimenti compiuti da Joaquin Errazuriz e da uno di noi (Harry P. Gregor), si è riusciti, per ultrafiltrazione, a innalzare la concentrazione dell'alga *Selenestrum* da 0,2 grammi per litro a 17 grammi per litro. In questo processo le cellule non subiscono rottura e il liquido filtrato risulta completamente limpido.

La somministrazione di sali nutritivi alle alghe è un problema più difficile da risolvere. John H. Ryther e Joel Goldman, della Woods Hole Oceanographic Institution, hanno calcolato che, se l'intera produzione industriale statunitense di fertilizzanti fosse incanalata in questo tipo di coltivazione destinata a combustibile, ne sarebbe soddisfatto solo dall'1 al 3 per cento del fabbisogno. Tale fabbisogno può essere diminuito solo attraverso la conservazione dei sali nutritivi e, per questa pratica, si potrebbe far ricorso a numerose tecnologie basate sull'utilizzazione di membrane.

È attualmente in funzione un prototipo di sistema basato sulla coltivazione delle alghe, costruito da Daniel Omstead, Thomas Jeffries e uno di noi (Harry P. Gregor). Il materiale che si forma nel digestore batterico viene sottoposto a ultrafiltrazione per separarne i sottoprodotti, circostanza questa che presenta il vantaggio di accelerare sostanzialmente la digestione. Il prodotto finale della digestione è metano; come sottoprodotti si ottengono bicarbonato di ammonio, sali di acidi organici e comuni sali inorganici. Il liquido filtrato contenente queste sostanze viene fatto passare attraverso una cella a membrana, per la dissociazione dell'acqua, consentendo il recupero di ammoniaca e anidride carbonica. L'anidride carbonica viene disciolta in una soluzione basica e riciclata nel digestore per incrementare la produzione di metano; l'ammoniaca può essere riciclata oppure venduta come fertilizzante. Gli acidi organici, che hanno valore commerciale, vengono separati per estrazione con solvente. Si fa infine ricorso all'osmosi inversa e all'elettrodialisi per purificare i restanti sali nutritivi destinati a essere riciclati, per concentrare i residui da eliminare e per recuperare acqua pura. Naturalmente, anche impiegando questo elaborato sistema di cicli chiusi, una parte dei sali nutritivi va perduta. Una possibile soluzione per compensare tale deficit potrebbe essere quella di impiegare fanghi di effluenti biologici come fertilizzanti. Una volta che siano stati concentrati con un processo di ultrafiltrazione, il trasporto dei fanghi potrebbe essere effettuato in modo economicamente interessante.

Vale la pena di notare che tutte le tecnologie basate su membrane hanno modelli nella cellula vivente. Un singolo organo infatti, il rene, svolge attività paragonabili a tutte e sei le tecnologie a membrana passate qui in rassegna: ultrafiltrazione, osmosi inversa, elettrodialisi, dissociazione dell'acqua, estrazione con solvente e promozione di reazioni chimiche con enzimi legati a membrana. Le membrane sintetiche non si avvicinano ancora a quelle biologiche quanto a velocità, compattezza, selettività ed efficienza. I principi di funzionamento delle membrane sintetiche sono, comunque, ben chiariti, attualmente, e le loro prestazioni appaiono di livello sufficiente per lo sfruttamento commerciale.

Il kivi

In Nuova Zelanda non vi sono stati mammiferi per 80 milioni di anni. Questo uccello ha potuto così colmare le nicchie ecologiche proprie di questi ultimi acquisendo alcune loro caratteristiche

di William A. Calder III

Il regno animale è così ricco di diversità che l'evoluzione può sembrare un processo casuale, come nel caso del kivi il noto uccello incapace di volare della Nuova Zelanda. Effettivamente, la selezione naturale è in parte occasionale, determinata com'è da mutazioni casuali e da circostanze storiche di discendenza e di localizzazione. Nel contempo, però, è sottoposta a precise limitazioni. Per esempio, sono le leggi fisiche a fissare quanto grande può essere un uccello per poter ancora volare. Nel breve arco della vita umana, è improbabile essere testimoni di interazioni tra ambiente ed eredità che diano origine a un organismo adattato a una nicchia ambientale disponibile. La storia naturale dell'isolamento biologico offre, però, alcuni esempi tipici, tra cui quello del kivi è certamente uno dei più affascinanti.

Si sente spesso dire che gli uccelli esibiscono adattamenti evolutivi meravigliosi. Si possono citare alcuni casi come la lunga migrazione della *Sterna paradisaea,* con le sue prodezze di navigazione; l'apertura alare di quei maestri del volo librato che sono i grandi albatri *(Diomedea exulans);* la coevoluzione dei minuscoli colibrì e delle piante che hanno fiori rossi tubolari e l'elaborato comportamento riproduttivo degli uccelli del paradiso. Si può dire la stessa cosa per un uccello che ha perso la capacità di volare, che esplora il terreno alla ricerca di vermi, che scava una tana nel fango per depositarvi un uovo ridicolmente più grande del normale e che è noto ai non biologi soprattutto per essere il marchio di un lucido da scarpe? La mia risposta è senz'altro positiva, considerato il contesto dell'isolamento e gli antenati da cui ha avuto origine.

Vi sono tre specie di kivi. Il kivi comune o bruno *(Apteryx australis)* si trova in Nuova Zelanda, nell'Isola del Nord, nell'Isola del Sud e anche nell'isola Stewart, sotto la punta meridionale dell'Isola del Sud. Il piccolo kivi maculato *(Apteryx oweni)* si trovava fino a poco tempo fa sia nell'Isola del Nord sia nell'Isola del Sud, ma ora si trova solo in quest'ultima. Anche il grande kivi maculato *(Apteryx haasti)* si trova solo nell'Isola del Sud. Questi uccelli non si trovano in nessuna altra parte del mondo.

I kivi rappresentano i membri più piccoli delle sette famiglie del gruppo dei ratiti (dal latino *ratis,* zattera). Questo gruppo comprende uccelli incapaci al volo che mancano dello sterno carenato a cui sono attaccati i robusti muscoli pettorali che fanno battere le ali degli uccelli che volano. Tra i ratiti vi sono anche gli struzzi, gli emù e i grandi moa giganti, ora estinti, della Nuova Zelanda. Oltre a mancare di sterno carenato, il kivi ha in comune con gli altri ratiti le caratteristiche del piumaggio che sono simili, nel kivi adulto, a quelle degli altri uccelli giovani. Le abbondanti piume del kivi sono lunghe e flessibili e non intrecciate tra loro e, quindi, il piumaggio del kivi ha un aspetto lanoso simile a quello di una pelliccia. Le piume non si scompongono quando il kivi scava la tana dove passa le ore diurne e gli forniscono un buon isolamento sia nel freddo e umido habitat sotterraneo sia in tutti gli altri ambienti in cui vive, che vanno dalle zone semitropicali dell'estremità settentrionale dell'Isola del Nord alle zone praticamente subpolari delle Alpi meridionali dell'Isola del Sud.

Il kivi non ha coda e le sue ali sono ridotte a moncherini nascosti sotto le piume. La sua sensibilità è aumentata da setole ben sviluppate presenti alla base del becco e da recettori sensitivi alla sua estremità. Ha zampe molto robuste che gli servono non solo per la locomozione ma anche per scavare e combattere. La posizione del kivi quando scava assomiglia a quella di un'anitra che nuota.

Il kivi si ciba di insetti, bacche, semi e altro materiale vegetale, ma il suo cibo principale è costituito dai lombrichi. La Nuova Zelanda ha una fauna molto ricca di lombrichi e il kivi ne ha a disposizione 178 specie originarie e 14 specie introdotte. La tecnica adottata dal kivi per cibarsi consiste nel sondare il terreno con il lungo becco e nello scovare il verme con il fiuto. Betsy G. Bang della Johns Hopkins University sostiene che, relativamente alla dimensione del prosencefalo, il bulbo olfattorio del kivi è il secondo in ordine di grandezza tra quelli di tutti gli uccelli che ha studiato. Nel kivi, il rapporto tra il diametro del bulbo olfattorio e il diametro dell'emisfero cerebrale è pari al 34 per cento. (Il rapporto più alto è quello della procellaria artica, o fumaro glaciale, che è pari al 37 per cento.) Questa constatazione è, rispetto alla funzione, di secondaria importanza, ma è convalidata dagli esperimenti condotti da Bernice M. Wenzel dell'Università della California a Los Angeles. Essa ha messo a disposizione di alcuni kivi dei vasi coperti da una rete e pieni di terra, nel fondo dei quali vi erano dei recipienti, alcuni contenenti cibo, altri solamente terra. Se il recipiente era pieno solo di terra, i kivi ignoravano il vaso e non provavano neppure a cercare nel suo interno.

La biologia del kivi appare nella sua luce migliore se viene esaminata nel contesto di aspetti apparentemente diversi: l'ambiente naturale della Nuova Zelanda, l'evoluzione complessiva dei ratiti e la funzione delle uova d'uccello e dei loro gusci. La topografia della Nuova Zelanda è spettacolarmente erta, con elevazioni dovute alla formazione di faglie e al vulcanismo e modellate dall'erosione glaciale. La pioggia è generalmente abbondante e prima che i coloni europei cominciassero ad alterare il paesaggio, predominava la foresta. Alle altitudini e alle latitudini inferiori le foreste erano un misto di latifoglie e di podocarpi, un gruppo di conifere caratteristico del Pacifico sudoccidentale e del Cile. Sopra questa fascia di foresta mista e su fino al limite della vegetazione arborea, tra i 900 e i 1500 metri, nella foresta predominavano i faggi meridionali. Erano anche presenti distese erbose naturali, soprattutto sui lati più asciutti sottovento delle catene montuose principali come l'area a est delle Alpi meridionali.

Tranne due specie di pipistrelli, nella Nuova Zelanda non vi erano mammiferi finché arrivò l'uomo nel primo millennio dopo Cristo. L'evoluzione delle due famiglie di ratiti - i kivi e i moa - poté quindi procedere libera da predazione e competizione, comuni negli ambienti abitati da mammiferi terrestri. È stato qui che i ratiti hanno avuto il loro massimo successo nella diversificazione. W. R. B. Oliver del Canterbury Museum di Cristchurch, uno studioso dei moa recentemente estintisi, ha attribuito i loro resti semifossilizzati a 22 specie diverse. In base a una recente analisi, Joel Cracraft del Medical Center dell'Università dell'Illinois sostiene che i fossili rappresentano solo 13 specie. Ciononostante, insieme alle tre specie di kivi viventi, si tratta di un totale di 16 specie di ratiti distribuite su un'area che è solo la metà del Colorado, anche se copre un arco di latitudini uguale a quello compreso tra Daytona Beach, in Florida, e Goose Bay, in Terranova.

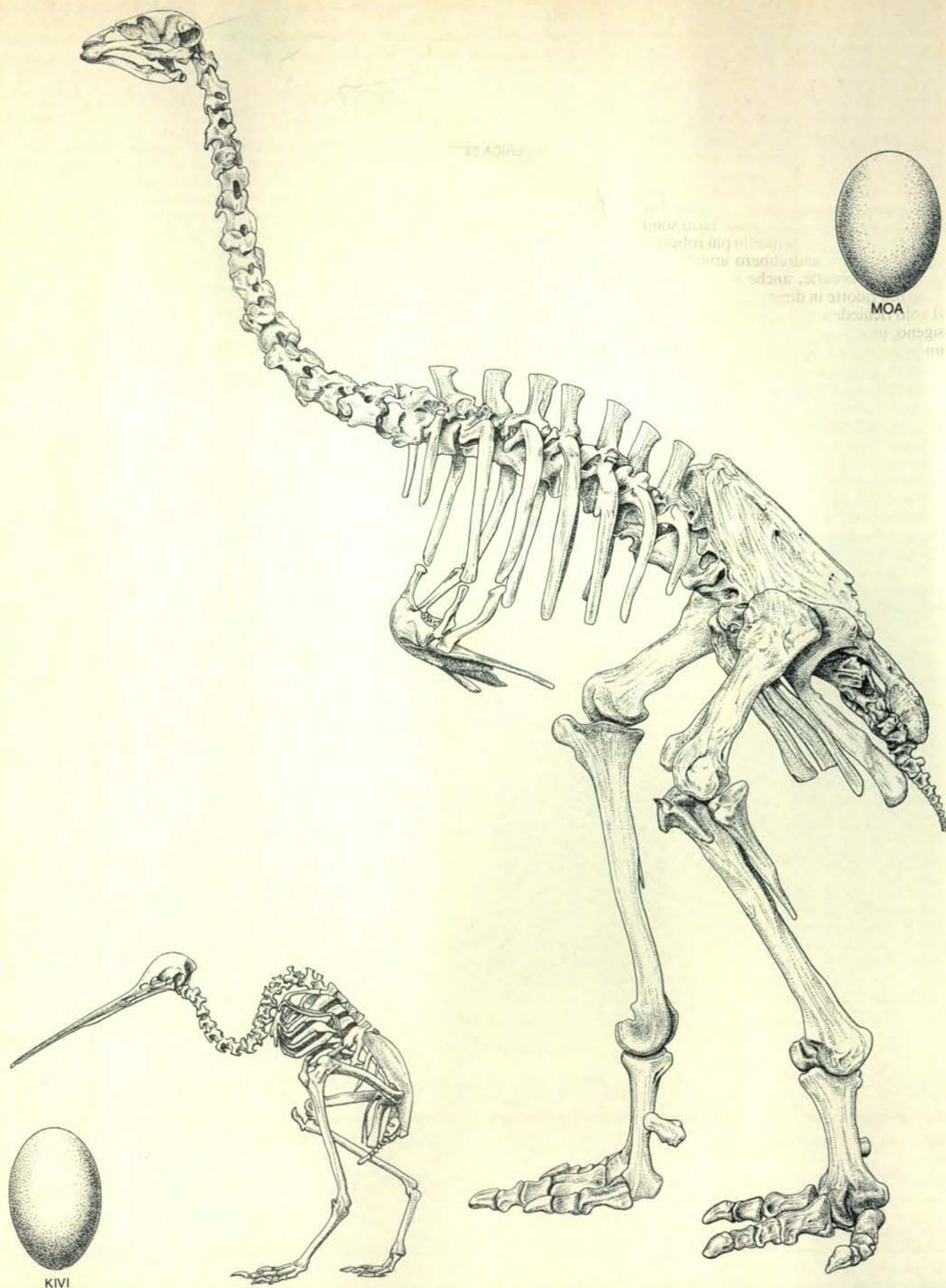
Alcuni dei moa evolutisi in isolamento su isole da cui erano incapaci di allontanarsi erano alti, da adulti, solo mezzo metro; altri arrivavano fino a tre metri. Dean Amadon dell'American Museum of Natural History valuta che le due razze della specie di maggiori dimensioni, *Dinornis giganteus,* pesassero tra i 230 e i 240 chilogrammi, più del doppio del peso di uno struzzo. Un kivi bruno adulto non è più alto di una gallina e pesa in media 2,2 chilogrammi.

I primi mammiferi, se si eccettuano i pipistrelli, a invadere la Nuova Zelanda sono stati gli uomini: un gruppo di navigatori polinesiani, noti attualmente come maori o cacciatori di moa, che raggiunsero le isole nel IX secolo dopo Cristo. Se portassero con loro cani e ratti, non si sa; lo fecero tuttavia i polinesiani che arrivarono successivamente.

Poiché, fino all'arrivo dell'uomo, i moa erano in Nuova Zelanda gli animali terrestri di maggiori dimensioni, costituivano ovviamente una preda per gli invasori. Sebbene alcune specie di moa probabilmente siano scomparse ancora prima dell'arrivo dell'uomo, quelle che sopravvissero furono probabilmente sterminate prima dell'arrivo degli europei, con lo sbarco del capitano Cook nel 1769. La grande «invasione» dei mammiferi cominciò all'inizio del XIX secolo, quando i coloni europei introdussero in Nuova Zelanda 22 diversi generi di mammiferi, comprendenti suini, ovini e bovini. Molti di questi «immigranti» furono introdotti intenzionalmente come selvaggina oppure si sottrassero alla domesticazione e parecchi di questi cominciarono presto a provocare la distruzione ecologica.

Dal momento del loro arrivo gli europei hanno esercitato sulla flora e sulla fauna insulari pressioni di ogni genere. L'habitat forestale del kivi fu sottoposto a una duplice aggressione: l'abbattimento degli alberi originari e la creazione di radure per i pascoli. Negli anni recenti molte estensioni forestali originarie sono state rimpiazzate con il pino di Monterey, che venne importato data la maggiore rapidità di crescita rispetto alle specie originarie. Inoltre si ricorse a trappole e veleni nel tentativo di controllare i mammiferi importati che erano diventati nocivi. In base a quello che attualmente si sa sulla persistenza dei veleni nelle catene alimentari riscontrata altrove, sembra probabile che i veleni abbiano influito negativamente sugli animali originari, compreso il kivi. In effetti, F. C. Kinsky del Dominion Museum di Wellington ha riferito di casi di kivi uccisi dal veleno. I kivi sono minacciati anche dalle trappole, dalle automobili, dalla predazione dei cani e dal fuoco con cui si distruggono lembi di foreste.

I kivi ora sono protetti, eppure il piccolo kivi maculato sta scomparendo. Brian Reid e Gordon Williams del Wildlife Service della Nuova Zelanda pensano che l'uccello sia scomparso dall'Isola del Nord prima che gli europei si stabilissero nell'isola. Non sembra probabile che sia stato cacciato dai polinesiani fino al punto di estinguersi, perché il kivi bruno, che è ugualmente vulnerabile nei confronti della caccia, è ancora presente nell'Isola del Nord. A complicare ulteriormente l'enigma, sembra che il processo continui. Reid riferisce che, anche se, nell'Isola del Sud, il piccolo kivi maculato non è ancora scomparso, si vede molto raramente.

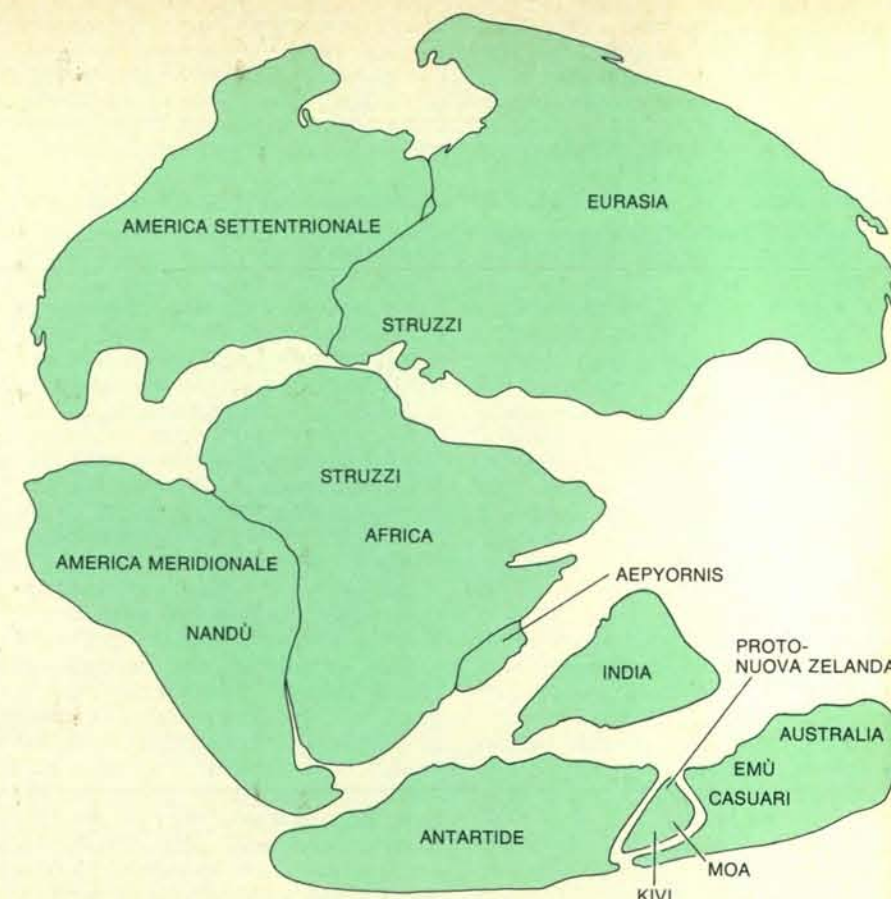
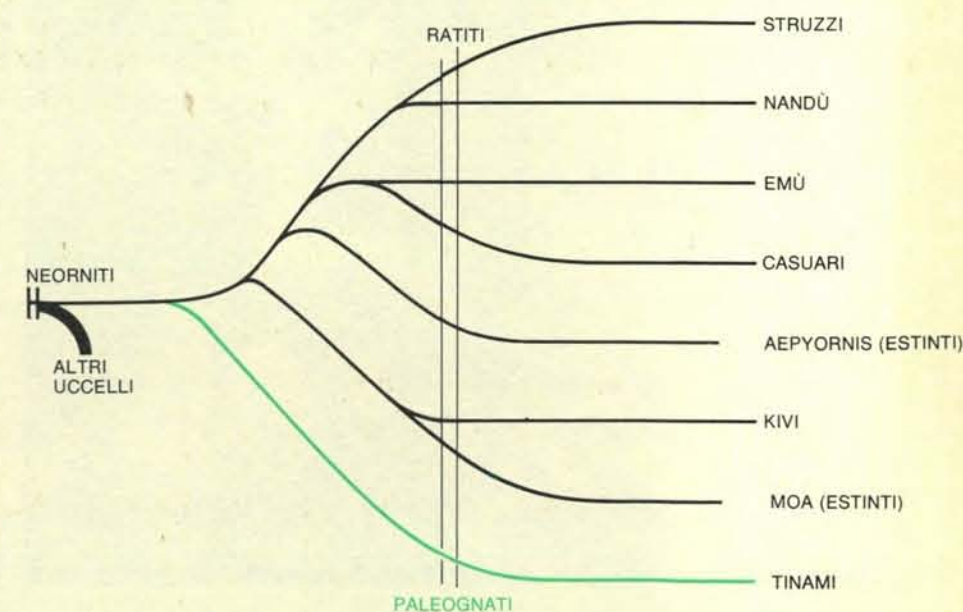
Il kivi comune, o kivi bruno *(Apteryx australis),* è la principale delle tre specie di kivi presenti in Nuova Zelanda. Ha un aspetto lanoso in quanto il suo piumaggio assomiglia più al piumino che alle penne degli uccelli. Il kivi si ciba in prevalenza cercando vermi col lungo becco.

Nell'illustrazione si vedono di profilo gli scheletri di un kivi e di un moa, un parente gigante ora estinto del kivi. Gli esemplari sono alti rispettivamente 30 e 90 centimetri. La specie di moa è infatti, in questo caso, una delle più piccole. Membri dei ratiti, uccelli incapaci di volare, sottoordine degli uccelli, i kivi e i moa sono anatomicamente simili, con uno sterno piatto, ali rudimentali e forti zampe. Le uova sono circa delle stesse dimensioni; l'uovo di moa è leggermente superiore in diametro e quindi avrebbe pesato circa il 35 per cento in più di un uovo di kivi.

I kivi e i moa, a essi strettamente imparentati, hanno parenti in tutto il mondo. Oltre allo struzzo dell'Africa e l'emù dell'Australia, i ratiti includono i casuari dell'Australia e della Nuova Guinea, i nandù dell'America meridionale e l'*Aepyornis*, un uccello estinto originario del Madagascar e delle Canarie. Come si sono evoluti tutti questi uccelli incapaci di volare?

Dall'anatomia dei ratiti si capisce che i loro antenati erano abili al volo. Per esempio, le ossa del carpo dei ratiti sono saldate tra loro, per renderlo più robusto, nel punto dove andrebbero attaccate le remiganti primarie, anche se le loro ali sono ora ridotte in dimensioni e funzioni. Il volo richiede un elevato apporto di ossigeno, un supporto meccanico robusto e un preciso coordinamento; la fisiologia dei ratiti mostra ancora che i loro antenati facevano fronte a queste esigenze. Gli uccelli hanno conservato il sistema di sacchi aerei per un'efficace ventilazione dei polmoni, una fusione delle vertebre sacrali con le ossa del bacino e un accorciamento delle vertebre codali che serve a irrobustirle, così come un grande cervelletto, necessario al coordinamento dei movimenti.

È opinione comune che tutti gli uccelli abbiano avuto un antenato comune delle dimensioni di un colombo che viveva nel Giurassico, 150 milioni di anni fa, e quindi è ragionevole chiedersi quali fattori abbiano favorito la selezione naturale di uccelli delle dimensioni dei ratiti. I vantaggi di un corpo di grosse dimensioni comprendono una maggiore velocità nella corsa, una maggiore possibilità di estendere il proprio territorio e una dominanza potenziale maggiore sugli altri animali. Un altro vantaggio è costituito da una maggiore resistenza nei periodi in cui il cibo è scarso, dato che la quantità di energia che un animale può immagazzinare è direttamente proporzionale alla sua massa mentre il metabolismo, ossia la velocità a cui l'energia immagazzinata viene consumata, non aumenta con l'aumentare della massa.

Allo stesso tempo, però, un corpo più grosso richiede una maggiore quantità totale di cibo e questo è potenzialmente uno svantaggio. Forse uno svantaggio ancora maggiore è costituito dal fatto che un uccello di grandi dimensioni può pesare troppo per volare. C. J. Pennycuick dell'University College di Nairobi ha messo in evidenza come l'energia necessaria a un uccello per stare sollevato in aria aumenti molto più rapidamente dell'energia che i suoi muscoli pettorali possono sviluppare, al crescere della sua corporatura. Un uccello che pesi più di 12 chilogrammi non è in grado di sollevarsi da terra. Gli uccelli giganti hanno perso, quindi, un vantaggio evolutivo in cambio di un altro.

Su un altro enigma evolutivo si è discusso molto fino a poco tempo fa. Si tratta del perché i ratiti che non volano abbiano avuto una diffusione geografica così vasta. I ratiti erano una volta considerati uccelli primitivi sia per la loro incapa-

Si può spiegare l'attuale distribuzione dei ratiti supponendo che essi abbiano avuto origine da un comune antenato che ha perso la capacità di volare prima che la deriva dei continenti dividesse l'antico continente di Gondwana nei continenti attuali dell'emisfero australe. Secondo Joel Cracraft del Medical Center dell'Università dell'Illinois, i kivi e i moa sono anatomicamente più primitivi degli altri ratiti. Il loro isolamento può essere dovuto alla separazione della Nuova Zelanda, come è mostrato qui, dall'Antartide occidentale 80 milioni di anni fa. L'*Aepyornis* del Madagascar ebbe origine dal ceppo originario rimasto, così come i casuari e gli emù dell'Australia e della Nuova Guinea, gli struzzi dell'Africa e dell'Eurasia e i nandù dell'America meridionale.

L'albero genealogico dei ratiti, basato sugli studi anatomici di Cracraft, pone le sette famiglie su uno stesso ramo con i tinami (o cripturi), uccelli sudamericani analoghi alle pernici che hanno conservato la capacità di volare *(in colore)*. Per Cracraft struzzi e nandù sono strettamente imparentati. Studi immunologici suggeriscono, però, che i nandù siano molto più vicini ai kivi.

cità di volare, sia perché sono separati gli uni dagli altri dagli oceani che essi non sono in grado di attraversare. Sembrava necessario, allora, formulare ipotesi che spiegassero l'evoluzione indipendente di ciascun membro del gruppo. Il biologo T. H. Huxley era pressoché il solo a opporsi a tale ipotesi di un'origine proveniente da più *phila*. Egli basava la sua opposizione su studi condotti sul palato osseo degli uccelli.

Attualmente gli studi anatomici di Cracraft e di Walter J. Bock della Columbia University (riguardanti rispettivamente lo scheletro postcraniale degli uccelli e il loro cranio) indicano che i ratiti fanno parte del gruppo di uccelli più evoluti e non di quelli primitivi. La più significativa tra le loro caratteristiche distintive è la morfologia del palato analoga a quella di un uccello, simile a una pernice, appartenente all'ordine dei cripturi, originario dell'America meridionale. Si può quindi supporre che i cripturi e i ratiti abbiano antenati comuni e che l'allontanamento dei ratiti dall'albero genealogico degli uccelli capaci di volare non sia stato anteriore all'origine dei cripturi.

Certe proprietà delle proteine presenti nell'albume delle uova dei ratiti e le loro caratteristiche immunologiche sono state analizzate da David Osuga, Robert E. Feeney, Ellen Prager e Allan C. Wilson dell'Università della California a Davis e dell'Università della California a Berkeley. I risultati da loro ottenuti hanno consentito di costruire un albero genealogico dei ratiti molto simile a quello proposto da Cracraft. Tra i due alberi, la sola differenza significativa consiste nel fatto che Cracraft imparenta strettamente gli struzzi ai nandù sulla base delle caratteristiche anatomiche. Altri hanno messo in rilievo come i due uccelli hanno pidocchi dello stesso genere e che mentre i ratiti del Pacifico hanno solo ali rudimentali, gli struzzi e i nandù hanno ancora ali abbastanza sviluppate che usano nel corteggiamento. Gli studi immunologici, però, suggeriscono che i nandù siano effettivamente più vicini ai kivi che non agli struzzi. Sembra che queste scoperte, a parte alcuni piccoli punti ancora da definire, abbiano affossato le ipotesi di un'origine polifiletica, supponendo, naturalmente, che la strana distribuzione geografica dei ratiti possa essere soddisfacentemente spiegata in altro modo.

Cracraft ha individuato una spiegazione semplice nella deriva dei continenti. Il modello di distribuzione dei ratiti corrisponde esattamente alla frattura del grande continente meridionale di Gondwana che ebbe inizio nel tardo Giurassico. Alla fine del Cretacico, circa 90 milioni di anni fa, l'Africa venne completamente divisa dall'America meridionale, fornendo così la possibilità di un primo isolamento dei nandù e degli struzzi che si adatta alla distanza immunologica tra i due. La Nuova Zelanda cominciò ad allontanarsi dall'Antartide occidentale circa 80 milioni di anni fa, portando all'isolamento l'antenato ratite, qualunque esso fosse, che ha dato origine ai moa e ai kivi.

In questa microfotografia elettronica a scansione, con un ingrandimento di 325 volte, si vede l'interno di un uovo di kivi. I cristalli di calcite secreti dalla ghiandola preposta alla produzione del guscio formano una serie di depositi grosso modo circolari: i coni mammellari. La secrezione continuata fa sì che i coni crescano, e gli interstizi che rimangono costituiscono i pori respiratori del guscio. La microfotografia è stata scattata da Cynthia Carey dell'Università del Colorado.

Quando ebbero origine le diverse dimensioni dei ratiti della Nuova Zelanda? Il kivi è forse un moa rimpicciolito? Gli unici resti noti di moa sono relativamente recenti, così nulla si sa sul ceppo originario dei ratiti. Se si accetta l'ipotesi monofiletica dell'origine dei ratiti e la probabilità che la capacità degli uccelli al volo si sia persa a mano a mano che le loro dimensioni aumentavano, si è portati a concludere che i kivi siano considerevolmente più piccoli dei loro antenati.

Questi studi rientrano nell'allometria, che Stephen J. Gould della Harvard University ha accuratamente definito come «lo studio della dimensione e delle sue conseguenze». Per esempio, se la dimensione di un organismo è aumentata, non tutte le sue caratteristiche anatomiche possono cambiare nello stesso rapporto se devono conservare la stessa funzione. Uno scheletro o un guscio d'uovo deve essere sproporzionatamente più pesante in un organismo più grande per fornire un margine di sicurezza sotto il peso di una massa maggiore. D'altra parte, se la massa di un organismo è doppia, il suo tasso metabolico non è raddoppiato, ma aumenta solo del 68 per cento e le frequenze cardiache e respiratorie diminuiscono del 20 per cento circa.

Le analisi allometriche sono estremamente semplici, perché entro una determinata classe di organismi non solo le dimensioni strutturali ma anche i tassi e le frequenze dei processi organici sono funzioni esponenziali del peso corporeo. Quando le funzioni sono riportate su scale logaritmiche, mostrano generalmente tendenze rettilinee.

Il primo studio allometrico del rapporto tra la massa delle uova d'uccello e la massa dell'uccello che le depone fu pubblicato da Julian Huxley, nipote di T. H. Huxley. Per esempio un colibrì che pesa 3,5 grammi depone un uovo di 0,5 grammi e uno struzzo che pesa 100 chilogrammi depone un uovo il cui peso varia da 1,4 e 1,7 chilogrammi. In altre parole, sebbene l'uovo di uno struzzo sia molto più pesante di quello del colibrì, è molto leggero rispetto al peso dello struzzo. Ricorrendo a una interpolazione allometrica, un kivi che abbia le dimensioni di una gallina dovrebbe deporre un uovo grande come quello della gallina o forse leggermente più grande, con un peso che va dai 55 ai 100 grammi. Invece l'uovo di un kivi bruno pesa generalmente dai 400 ai 435 grammi. Quanto dovrebbe pesare un uccello che depone un uovo di 400 grammi? Secondo l'equazione allometrica, 12,7 chilogrammi, circa sei volte il peso del kivi.

I maggiori esperti americani di analisi allometrica sulle uova degli uccelli sono Hermann Rahn e i suoi colleghi della State University di New York a Buffalo: Amos Ar, Charles V. Paganelli, Robert B. Reeves, Douglas Wangensteen e

Donald Wilson. Il loro lavoro ha permesso di ottenere equazioni che, partendo dal peso dell'uovo, prevedono un'enorme quantità di proprietà correlate: lo spessore del guscio, la permeabilità del guscio ai gas e al vapore acqueo, il periodo probabile di incubazione, il tasso massimo di assunzione di ossigeno e la perdita totale d'acqua durante l'incubazione. Per esempio, la diffusione del vapore acqueo dall'albume attraverso i pori del guscio è misurabile come perdita di peso durante il periodo di incubazione. Il tasso di perdita può determinare la conduttanza di vapore acqueo (permeabilità) nello stesso modo in cui si determina la conduttività elettrica o la resistenza una volta noti la corrente e la tensione di un circuito elettrico. L'ossigeno diffonde verso l'interno nello stesso modo, ossia attraverso gli stessi pori, così che la massima assunzione di ossigeno è direttamente proporzionale alla permeabilità, ossia alla conduttanza di vapore acqueo. Quando si conosce, quindi, il tasso di perdita acquea, si può prevedere il metabolismo dell'embrione fino al momento della schiusa.

Lo sviluppo dell'uovo di uccello ha inizio quando si forma nell'ovaio una cellula uovo consistente di una parte germinativa e di una parte di deutoplasma, o tuorlo, che serve da sostanza nutritiva per il futuro embrione. L'albume, le membrane, il guscio si aggiungono dopo l'ovulazione, quando la cellula uovo scende nell'ovidotto. Il rapporto tra il tuorlo e la massa totale dell'uovo è direttamente proporzionale alla durata dello sviluppo embrionale prima della schiusa. Gli uccelli inetti, cioè quelli che escono dal guscio ciechi e senza piume, hanno uova con forse il 20 per cento di tuorlo. Gli uccelli precoci, che escono dal guscio già coperti di piumino e capaci di nutrirsi da soli, provengono da uova con il 35 per cento di tuorlo. Considerata la notevole dimensione dell'uovo di kivi e il lungo periodo di incubazione (da 71 a 75 giorni invece dei 44 giorni che sarebbero normali per un uovo di 400 grammi) non sorprende il fatto che il piccolo kivi nasca completamente coperto di piume e che si cibi da solo. Reid ha trovato che il contenuto delle uova di kivi è costituito per il 61 per cento da tuorlo , una quantità proporzionalmente doppia rispetto a quella contenuta nelle uova di uccelli precoci tipici.

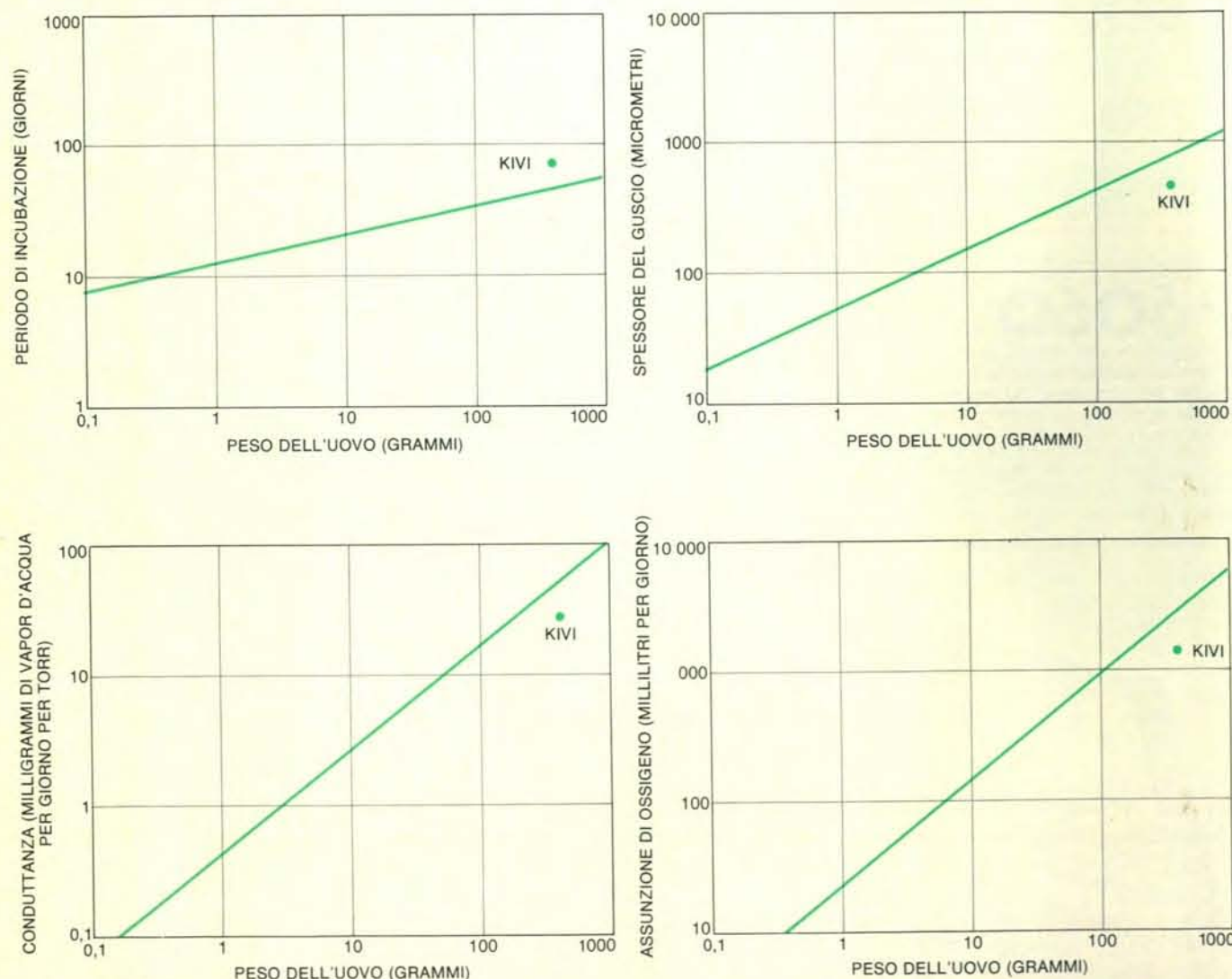
Una volta che il guscio dell'uovo si sia formato non è più possibile aggiungere sostanze nutritive o acqua e, quindi, il contenuto dell'uovo deve comprendere abbastanza tuorlo per nutrire l'embrione. Deve includere anche acqua a sufficienza e l'integrità della struttura del guscio deve essere tale da conservare l'acqua in maniera efficiente. Rudolph Drent, che si trova ora all'Università di Groninga in Olanda, ha misurato la perdita giornaliera d'acqua da parte di uova piccole e grosse durante l'incubazione naturale. Partendo dai risultati di Drent e tenendo in considerazione il rapporto tra durata dell'incubazione e dimensioni dell'uovo, Rahn e Ar hanno stabilito che la percentuale di perdita d'acqua è quasi la stessa in tutte le uova: il 18 per cento per le uova più piccole e il 13 per cento per quelle più grosse.

Correlazioni attendibili come queste indicano in maniera evidente quanto la selezione naturale operi entro limiti fisici molto stretti che negli uccelli, in particolare, impongono certi criteri nelle dimensioni dell'uovo, nel periodo di incubazione e in alcuni dettagli strutturali come la porosità del guscio. Se l'evoluzione dell'uovo negli uccelli si è svolta entro, si fa per dire, rigorose tolleranze tecniche, che cosa dire del kivi? Le regole che sovraintendono il processo evolutivo sono realmente conosciute? Le regole possono essere vere anche se alcuni uccelli ne sono esenti? È stato ponendomi questi interrogativi che ho deciso recentemente di trascorrere un anno sabbatico in Nuova Zelanda a studiare i kivi.

Uno degli interrogativi che mi si presentavano più frequentemente nelle 14 ore di volo da Los Angeles ad Auckland riguardava il periodo estremamente lungo di incubazione del kivi. La lunga durata dell'incubazione poteva dipendere da un contatto fisico insufficiente per il trasferimento del calore da un uccello relativamente piccolo a un uovo relativamente grosso? Non vedevo l'ora di controllare le temperature dell'uovo.

Mi ero seduto a un ristorante prospiciente il lago Taupo, in Nuova Zelanda, per prendere un tè, quando la mia attenzione fu attratta da una ripresa televisiva, in cui veniva intervistato Barry Rowe della Società zoologica di Otorohanga a proposito delle sue ricerche sul kivi. Guardavo, del tutto incredulo, Rowe che apriva la «porta» di una tana artificiale di kivi e rimuoveva un uovo che apriva in due metà, rivelando la presenza all'interno di un piccolo trasmettitore che forniva in continuazione la temperatura dell'uovo. Dopo essermi ripreso dallo shock di apprendere che un altro ricercatore mi aveva battuto sul tempo, gli telefonai ed egli mi invitò a dividere con lui le apparecchiature e l'ospitalità della società zoologica. Cominciai subito a lavorare con Rowe e presto mi trovai impegnato a misurare le temperature di uova rinforzate, piene d'acqua, mentre venivano incubate.

Rowe registrava le variazioni di peso della femmina di kivi durante l'intervallo tra due deposizioni successive. L'accumulazione di materiale che viene depositato nel tuorlo ricco di energia occupa i primi sette giorni e mezzo del ciclo ovulatorio di 24 giorni. In termini di investimento energetico la femmina, durante questa settimana, deve aumentare il suo metabolismo basale dal 174 al 203 per cento. L'attività di alimentazione dell'uccello allo stato libero aumenta probabilmente in maniera notevole nello stesso periodo, ma non vi sono informazioni di tipo quantitativo sul comportamento dei kivi in natura che confermino questa supposizione.

Il kivi femmina, dopo aver investito una grande quantità di energia nei 24 giorni di preparazione dell'uovo, cessa di cibarsi negli ultimi due giorni del ciclo. Quindi, quando lascia la tana dopo aver deposto l'uovo, ha perso più peso del peso

Le relazioni tra le uova di uccello sono evidenti in questi grafici rettilinei riportati su scala logaritmica che mettono a confronto l'aumento di peso dell'uovo (scala orizzontale) con quattro fattori determinati da Hermann Rahn e dai suoi collaboratori della State University di New York a Buffalo. Questi fattori sono l'aumento del periodo di incubazione (*in alto a sinistra*), l'aumento di spessore del guscio (*in alto a destra*), l'aumento del tasso di evaporazione dall'uovo (*in basso a sinistra*) e l'aumento di assunzione di ossigeno da parte dell'uovo (*in basso a destra*). In ognuno di questi casi l'uovo di kivi risulta anomalo. Il suo periodo di incubazione è più lungo del 60 per cento circa di quello di un uovo delle stesse dimensioni, il suo guscio è più sottile del 40 per cento e il tasso di diffusione del vapor d'acqua e quello di assunzione di ossigeno sono al di sotto del normale anche se i pori più piccoli di un guscio più sottile dovrebbero accelerare il processo di diffusione.

Il peso di un uovo di uccello non è direttamente proporzionale al peso della femmina. Gli istogrammi del grafico in alto mostrano il peso dell'uovo in percentuale del peso del genitore per un piccolo uccello ipotetico che pesa solo 10 grammi, per un grosso uccello ipotetico che pesa 100 chilogrammi e per un kivi bruno che pesa 2,2 chilogrammi. Il peso dell'uovo del grosso uccello è meno del 2 per cento del peso del genitore. Gli istogrammi in basso mostrano la perdita di peso durante l'incubazione delle uova, espressa in percentuale del peso dell'uovo appena deposto. L'uovo piccolo richiederebbe un'incubazione di 13,4 giorni, quello grande di 62,3 giorni e l'uovo di kivi da 71 a 74 giorni. Le differenze percentuali nella perdita di peso non sono statisticamente significative. I calcoli si basano sulle formule messe a punto da Rahn e suoi collaboratori.

Nel kivi la formazione dell'uovo provoca nella femmina un aumento di peso di 400 grammi. Circa 200 grammi si aggiungono nei primi sette giorni del ciclo della durata di 24 giorni. Durante questa settimana le esigenze alimentari della femmina devono raggiungere il massimo. Circa 180 grammi di albume, membrana e guscio si formano mentre l'uovo scende nell'ovidotto. I minerali per il guscio sono forniti dalla dieta o prelevati dalle ossa. Negli ultimi due giorni del ciclo la femmina digiuna. Questi dati sono stati registrati da Barry Rowe della Società zoologica di Otorohanga.

dell'uovo stesso. L'investimento energetico nel futuro nato è ora completato. La femmina è libera di prendere cibo e può anche darsi che abbia già iniziato un secondo ciclo ovulatorio.

A questo punto il maschio si assume il compito di covare l'uovo per circa due mesi e mezzo. In cattività, lascia la tana per cibarsi una volta o due durante la notte. Alcuni osservatori hanno suggerito che allo stato libero la femmina porti occasionalmente cibo al maschio durante la cova, ma si sa di maschi che sono restati nella tana diversi giorni senza cibarsi. All'interno dell'uovo di kivi, dove si trova l'embrione, la temperatura di incubazione è in media di 35,4 gradi centigradi. Quando abbiamo posto un uovo contenente gli strumenti sotto un pollo bantam, un uccello più piccolo di un kivi, la temperatura di incubazione è salita a 37,7 gradi. La differenza riscontrata riflette la differenza della temperatura corporea dei due uccelli. Donald S. Farner dell'Università di Washington e i suoi colleghi, Norman Chivers e Thane Riney, hanno misurato le temperature corporee di varie specie di kivi e hanno trovato che la temperatura basale media del kivi bruno adulto è di 38 gradi centigradi, ossia inferiore di due gradi alla temperatura corporea di altri uccelli e più vicina a quella dei mammiferi. In ogni caso se il pollo bantam, che è più piccolo, può mantenere l'uovo di kivi a una temperatura di 37,7 gradi, il lungo periodo di incubazione non dovrebbe essere semplicemente dovuto al fatto che l'uovo è troppo grande per essere tenuto caldo.

Conoscendo la temperatura dell'incubazione naturale del kivi, Rowe è riuscito a portare a termine l'incubazione artificiale di un uovo di kivi. Durante il periodo di incubazione di 71 giorni ho provveduto a misurare l'assunzione di ossigeno dell'uovo a intervalli di due o tre giorni. Nel frattempo, poiché è difficile procurarsi uova di kivi, ci servimmo di uova sterili per altre misurazioni comparative. Il kivi si è rivelato «anticonformista» non solo nelle dimensioni dell'uovo, ma anche nella temperatura e nella durata dell'incubazione. L'uovo di kivi si è adeguato in qualche modo alle previsioni di Rahn e dei suoi collaboratori?

Io ho misurato la conduttanza (permeabilità) al vapore acqueo del guscio, il suo spessore e ho separato il tuorlo dall'albume, che sono stati portati al Centro di ricerche agricole di Ruakura, dove C. R. Parr, D. P. Karl e R. Whaanga hanno misurato il contenuto d'acqua di entrambi e hanno preparato campioni per la misurazione del contenuto energetico dell'uovo per mezzo di un calorimetro a bomba. Il contenuto energetico medio di cinque uova era di 959 chilocalorie. Il valore concordava con una precedente valutazione fatta da Reid, che aveva supposto un'analogia tra l'uovo di kivi e l'uovo del pollo da cortile.

Questa scoperta significa che in proporzione alle dimensioni e al tasso metabolico, il kivi è in testa a tutti gli uccelli per quanto riguarda la quantità di energia contenuta nell'uovo. Questa enorme riserva di energia sostiene l'embrione di kivi durante il suo lunghissimo sviluppo. Inoltre, quando il piccolo kivi emerge dall'uovo, è stato consumato solo poco più di metà del tuorlo. Ciò che rimane nel sacco del tuorlo viene assorbito dall'addome del pulcino e provvede a nutrirlo per le due successive settimane circa.

L'assunzione di ossigeno da parte dell'embrione di kivi durante il periodo di incubazione *(in colore)* è messa a confronto con l'assunzione prevista di ossigeno per un uovo normale di 400 grammi *(in nero)*. Il periodo di incubazione per l'uovo normale è di 45 giorni. Nel corso del primo mese di incubazione l'assunzione di ossigeno da parte dell'embrione di kivi aumenta rapidamente fino al secondo mese quando si stabilizza intorno a 60 millilitri per ora (1,4 litri per giorno). Le aree delle superfici delimitate dalle due curve sono uguali e questo dimostra che il costo totale metabolico dello sviluppo è circa lo stesso, indipendentemente dalla velocità dello sviluppo.

Che cosa dire della perdita d'acqua durante i 71-74 giorni di incubazione? Questo periodo è più lungo del 61 per cento rispetto alle previsioni allometriche circa il periodo richiesto per la schiusa di un uovo normale dello stesso peso. Con un'evaporazione continua attraverso i pori per un periodo così lungo, ci si aspetterebbe una perdita d'acqua superiore del 61 per cento, nel caso che gli altri fattori siano uguali, ma come abbiamo visto non lo sono.

La porosità del guscio dell'uovo di kivi è molto inferiore alla media. La riduzione dipende da una diminuzione del numero di pori e dalle loro dimensioni e, quindi, la permeabilità del guscio al vapore acqueo è solo il 60 per cento del normale valore previsto. Questa perdita di vapore acqueo al di sotto della norma compensa il periodo di incubazione anormalmente lungo. La diffusione dell'ossigeno verso l'interno dovrebbe naturalmente essere anch'essa più lenta, dato che avviene attraverso gli stessi pori. E di fatti è così, come risulta dalle misurazioni fatte: il tasso di assunzione di ossigeno è solo il 63 per cento del normale.

Sorge a questo punto un problema: l'assunzione più lenta di ossigeno da parte dell'embrione rappresenta una privazione parziale di ossigeno? Si è misurato il consumo di ossigeno di un piccolo kivi di due settimane che aveva finito tutto il contenuto del sacco del tuorlo, il consumo di ossigeno di due kivi un po' più grandi e anche di cinque kivi bruni adulti. Per i piccoli, il consumo medio di ossigeno a riposo era pari al 65 per cento del tasso previsto per uccelli del loro peso. La media per i kivi adulti era del 61 per cento del tasso metabolico previsto. Quando individui delle altre due specie di kivi furono analogamente analizzati, anche i loro tassi metabolici risultarono bassi. Di conseguenza, la ridotta porosità del guscio limita l'apporto di ossigeno durante l'incubazione grosso modo allo stesso livello che risulta normale per i kivi dopo la schiusa.

La base del nostro confronto era costituita dall'equazione allometrica per il metabolismo di tutti gli uccelli esclusi i passeriformi (l'ordine che comprende tutti gli uccelli canori). Potrebbe essere che anche gli altri ratiti abbiano un tasso metabolico inferiore alla media? In collaborazione con Terry Dawson dell'Università del New South Wales ho misurato il tasso metabolico degli emù e si è visto che anche loro hanno un tasso di consumo di ossigeno inferiore a quello previsto sulla base del peso corporeo. Forse il metabolismo basale di tutti i ratiti andrebbe riesaminato con una certa attenzione; può darsi infatti che sia al di sotto della media degli uccelli, come quello degli uccelli canori è al di sopra della media e come quello dei marsupiali è al di sotto della media dei mammiferi placentati

Proprio come la deposizione dell'uovo segna la fine dei doveri femminili, così il maschio termina il suo compito alla schiusa, a meno che la femmina non abbia deposto un altro uovo. Il piccolo kivi dopo la schiusa si nutre con il resto del tuorlo finché non è in grado di procurarsi il cibo da solo. Questo sistema di sostentamento funziona molto bene. Per esempio un kivi che alla nascita pesava 325 grammi scese a 225 prima di cominciare di nuovo a guadagnare peso nutrendosi da solo. La differenza in peso di 100 grammi corrispondeva al peso del tuorlo immagazzinato nell'addome.

Il rapporto tra il peso dell'uovo e il peso dell'adulto riscontrato nei kivi rappresenta in senso evolutivo una diminuzione delle dimensioni corporee o un aumento delle dimensioni dell'uovo? Per trovare una risposta, supponiamo che gli antenati dei kivi fossero uccelli abili al volo che arrivassero in Nuova Zelanda indipendentemente dai moa o da altri ratiti. In questo caso il grosso uovo dei kivi attuali piuttosto che rappresentare un legame con qualche antenato moa si sarebbe sviluppato per un processo di selezione naturale dopo l'arrivo di questi ipotetici pre-kivi abili al volo.

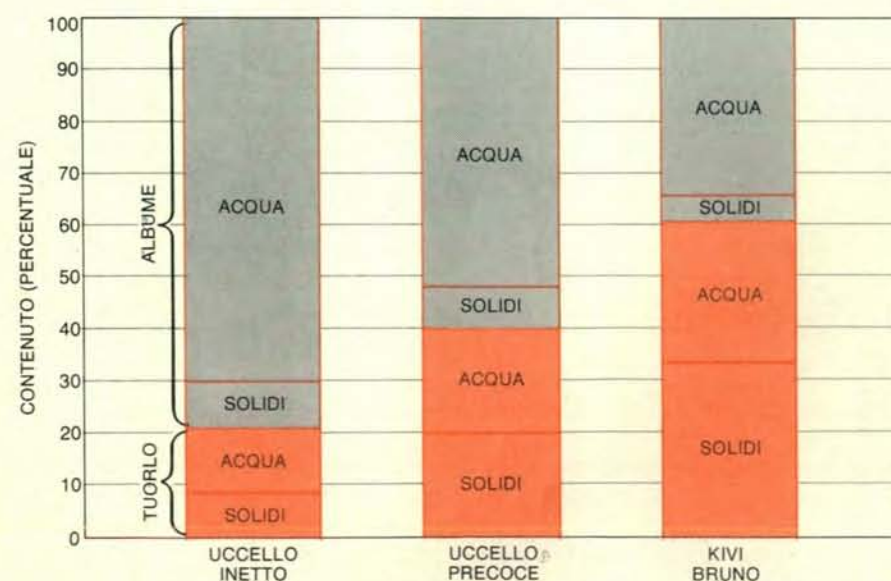
Quali sono i vantaggi selettivi di un uovo di grandi dimensioni? Favorirebbe uno sviluppo embrionale più lungo di quello tipico degli uccelli precoci, che potrebbe servire a programmare le richieste di energia in base alla naturale abbondanza di cibo. La disponibilità di cibo in Nuova Zelanda presenta però piccole fluttuazioni annuali e quindi non vi è alcuna necessità evidente di immagazzinare nell'uovo una grossa riserva di sostanze nutritive.

Parecchie specie di uccelli raggiunsero effettivamente la Nuova Zelanda in volo eppure hanno perduto o quasi la capacità di volare. Tra questi vi sono i weka, due specie di trampolieri incapaci di volare (*Gallirallus australis* e *Gallirallus hectori*); il takahe, un uccello acquatico incapace di volare della famiglia dei rallidi (*Notornis hochstetteri*); l'anitra incapace di volare dell'Isola Auckland (*Anas aucklandica*) e lo strigope, un passero che non vola più ma può ancora planare (*Strigops habroptilus*). Ognuno di questi uccelli, sebbene si siano evoluti in uno stato di incapacità al volo, hanno conservato la stessa proporzione delle dimensioni dell'uovo rispetto alle dimensioni corporee dei loro parenti che volano. Questo indica che non c'è in Nuova Zelanda alcun imperativo ambientale che metta in relazione l'incapacità al volo con le maggiori dimensioni dell'uovo. A questo proposito, le uova di moa, sebbene grandi, non lo sono in maniera così sproporzionata.

Per l'aumento nelle dimensioni dell'uovo deve esserci un vantaggio selettivo tanto forte da vincere qualsiasi forza che conservi le regole allometriche generali. Se, però, un tempo è esistita qualche pressione selettiva per una dimensione corporea minore, per esempio nell'evoluzione di un kivi da un antenato moa, la conservazione di un uovo sproporzionatamente grande potrebbe rappresentare semplicemente l'assenza di una forte pressione selettiva per un'economia delle dimensioni dell'uovo, favorita dal vantaggio di far nascere un kivi più sviluppato.

I pezzi della storia dei kivi possono essere cuciti in vari modi. Io preferisco considerare questo uccello curioso come un esempio classico di evoluzione convergente. Da questo punto di vista, un organismo d'uccello ha acquisito una notevole serie di caratteristiche che generalmente associamo ai mammiferi piuttosto che agli uccelli. Il fatto che l'arcipelago neozelandese con il suo clima temperato e la vasta estensione delle foreste fornisca habitat particolarmente adatti ai mammiferi è dimostrato dal successo dei mammiferi importati. Quando non c'erano mammiferi a reclamare le nicchie di questo ambiente ospitale, erano liberi di farlo gli uccelli.

I kivi depongono uova; dopo tutto sono uccelli, ma sono molto simili ai mammiferi sotto vari aspetti. Per esempio, Kinsky ha riferito che i kivi sono gli unici uccelli ad avere entrambi gli ovai pienamente funzionanti, così che le femmine hanno ovulazioni alternate come i mammiferi; come nei mammiferi, inoltre, il prolungato sviluppo embrionale del kivi avviene a una temperatura inferiore a quella degli uccelli. Il periodo di incubazione del kivi, da 70 a 74 giorni, è più vicino al periodo di gestazione di 80 giorni di un mammifero dello stesso peso che al periodo di 44 giorni necessario a far schiudere un uovo delle dimensioni di quello del kivi.

Quando, oltre queste cose, si considerano l'habitat sotterraneo (la tana) del kivi, il suo piumaggio lanoso che assomiglia a una pelliccia e il suo cibarsi notturno, per il quale sfrutta soprattutto l'odorato, le prove della convergenza sembrano sovrabbondare. Tra il serio e il faceto, vorrei aggiungere all'elenco il comportamento aggressivo del kivi. Durante le ricerche condotte alla Società zoologica di Otorohanga spesso ho dovuto penetrare in un grande recinto che era il territorio di un kivi maschio atto alla riproduzione. Quando mi intromettevo nel suo dominio di notte, correva verso di me soffiando come un gatto, mi afferrava le calze col becco e mi colpiva ripetutamente le caviglie con gli artigli finché me ne andavo via. Per questo comportamento e per le molte altre ragioni che ho detto mi sembra giusto attribuire a questo uccello la qualifica di mammifero onorario.

Viene messo a confronto il contenuto di tre uova. Sia per il tuorlo sia per l'albume il contenuto solido è distinto da quello acquoso. Lo stato di sviluppo di un piccolo al momento della schiusa dipende dalla percentuale di tuorlo che c'è nell'uovo. Il piccolo di un uccello cosiddetto «inetto» nasce nudo, cieco e inerme da un uovo che forse è per il 20 per cento tuorlo. Il piccolo precoce, incubato in un uovo con il 40 per cento di tuorlo, può correre dietro la madre un'ora dopo la nascita. Il piccolo del kivi bruno non riceve cure parentali. Nel periodo di incubazione è nutrito da un tuorlo che rappresenta il 61 per cento del contenuto dell'uovo; le parti solide del tuorlo sono per circa un terzo proteine e per due terzi grassi, quelle dell'albume sono in prevalenza proteine. I dati dell'uovo di kivi sono stati forniti da Brian Reid dello Wildlife Service della Nuova Zelanda.

GIOCHI MATEMATICI

di *Martin Gardner*

I versatili numeri di Bell: servono a calcolare il numero delle partizioni di un insieme, delle scomposizioni in fattori primi e persino dei possibili schemi di rima

Keeping time, time, time,
In a sort of Runic rhyme,
To the tintinnabulation that so
musically wells
From the bells, bells, bells, bells,
Bells, bells, bells—
From the jingling and the tinkling
of the bells.
EDGAR ALLAN POE, "The Bells"

S'immagini di avere cinque piatti, che contrassegnamo con *a, b, c, d* ed *e*, disposti in fila su un tavolo. Sul tavolo ci sono anche cinque pezzi degli scacchi: re, regina, alfiere, cavallo e torre. In quanti modi si possono disporre i pezzi sui piatti in modo che ci sia un solo pezzo su ogni piatto? La risposta è 5! Il punto esclamativo è il segno di fattoriale, il che indica che la risposta è $1 \times 2 \times 3 \times 4 \times 5$ ossia 120. Il problema, da un punto di vista combinatorio, è equivalente a quello di calcolare quante sono le possibili permutazioni di *a, b, c, d* ed *e*. In generale, per n oggetti le permutazioni sono $n!$

Si modifichino ora le regole così da permettere che su ogni piatto ci sia un numero qualsiasi di oggetti da uno a cinque. In quanti modi si possono disporre gli scacchi sui piatti? È ovvio che un pezzo può essere messo su un piatto in un modo solo, due pezzi su due piatti in quattro modi come si vede a sinistra nella figura della pagina a fronte. Se fate la prova con tre piatti e tre oggetti troverete che le combinazioni sono 27. Dato che 1, 4 e 27 sono uguali a 1^1, 2^2, 3^3 ci si può aspettare che n oggetti si possano mettere su n piatti in n^n modi differenti. E ciò si dimostra corretto: ad esempio, cinque scacchi si possono mettere su cinque piatti in 5^5 ossia 1325 modi differenti.

Se ci sono n oggetti e k piatti, allora, secondo le regole modificate, ci sono k modi di disporre gli oggetti sui piatti. Ad esempio si possono mettere due oggetti su tre piatti in 3^2, ossia nove, modi come si può vedere a destra nella figura della pagina a fronte. È facile scoprire il perché della formula: il primo oggetto si può disporre su k piatti in k modi diversi. Anche il secondo oggetto si può mettere su uno qualsiasi dei k piatti cosicché anch'esso si può disporre in k modi diversi. Dato che gli oggetti sono in numero di n, è chiaro che si possono disporre su k piatti in $k \times k \times k... \times k = k^n$ modi.

Consideriamo ora un problema più difficile. Abbiamo gli stessi cinque pezzi degli scacchi e gli stessi cinque piatti, ma ora i piatti non sono contraddistinti da lettere. In altre parole i piatti si considerano identici e quindi le loro posizioni sul tavolo sono inessenziali. Per esempio, se il re e la regina stanno su un piatto e gli altri tre scacchi su di un altro, non importa quale piatto contiene un gruppo e quale l'altro. Tutte le partizioni dell'insieme che pongono il re e la regina su un piatto e l'alfiere il cavallo e la torre su un altro, indipendentemente da quali due piatti vengano usati, verranno considerate identiche e conteranno come un solo modo di disporre i pezzi. Come si fa a calcolare tutti i modi in cui si possono disporre gli oggetti?

Ancora una volta, un oggetto si può mettere su un piatto in un modo solo, due oggetti si possono mettere su due piatti in due modi diversi: tutti e due su un piatto o uno su un piatto e uno su un altro. Questo caso è modello di molte situazioni reali. Per esempio, un marito e una moglie possono occupare due letti gemelli essenzialmente in due modi diversi: dormendo nello stesso letto o nei due letti separati. Un poliziotto può ammanettare due prigionieri in due modi diversi: può legare l'uno all'altro o ammanettarli separatamente. Nella figura in alto a pagina 110 si possono vedere i cinque modi in cui tre oggetti si possono disporre su tre piatti considerati identici. Questo caso rappresenta i cinque modi in cui tre persone possono occupare tre letti uguali, i cinque modi in cui tre nazioni possono formare alleanze e così via.

Come esperimento, potreste fermarvi a contare effettivamente i modi in cui quattro oggetti si possono disporre su quattro piatti considerati identici. In termini più tecnici, il problema è quello di determinare quante partizioni in sottoinsiemi non vuoti si possono fare su un insieme di quattro elementi distinti. Si troverà che sono precisamente quindici. Per un insieme di cinque oggetti il numero di partizioni sale a 52. Continuando ad aumentare il numero n di oggetti, si genera una successione di numeri: 1, 2, 5, 15, 52.... I numeri di questa successione, che sono estremamente utili nell'aritmetica combinatoria, si dicono numeri di Bell (che significa: campana) in onore del matematico americano di origine scozzese Eric Temple Bell, morto nel 1960. Essi sono strettamente legati ai numeri catalani che costituirono l'argomento del nostro articolo del novembre 1976.

Sebbene i numeri di Bell fossero stati scoperti già molto tempo prima che Bell ne scrivesse, egli fu il primo ad analizzarli in profondità e a mostrare la loro importanza. Nel suo primo articolo su tali numeri Bell spiegò come questi avessero destato il suo interesse. Egli aveva notato un errore su un manuale che dava il cosiddetto sviluppo in serie di Maclaurin per l'espressione e^{e^x} dove e è il numero trascendente di Eulero e x è un qualsiasi intero positivo. Lo sviluppo corretto è:

$$e\left(1+\frac{x}{1!}+\frac{2x^2}{2!}+\frac{5x^3}{3!}+\frac{15x^4}{4!}\cdots\right)$$

I coefficienti delle potenze di x sono precisamente i numeri di Bell. (Bell li chiamò numeri esponenziali, ma dopo che John Riordan li chiamò B in onore di Bell, divennero rapidamente noti come numeri di Bell.) Dallo sviluppo di Maclaurin è possibile derivare la cosiddetta formula di Dobinski che dà l'n-esimo numero di Bell, B :

$$B_n = \frac{1}{e}\sum_{k=0}^{\infty}\frac{k^n}{k!}$$

Bell era soprattutto un esperto di teoria dei numeri, ma è più conosciuto per il suo classico *The Development of Mathematics*, per l'opera più divulgativa *I grandi matematici* (Sansoni, Firenze, 1976) e altri libri per il grosso pubblico. I matematici più giovani rimarranno sorpresi nell'apprendere che negli anni venti e trenta Bell era un prolifico scrittore di fantascienza sotto lo pseudonimo di John Taine. Cinque dei suoi racconti sono stati ristampati in due volumi della Dover: *Seeds of Life and White Lily* e *The Time Stream*.

Torniamo ai numeri di Bell. I primi 13 compaiono a sinistra nella figura in basso di pagina 110. Per convenzione B_0 è uguale a 1. Come si può vedere, i numeri crescono con incremento esponenziale o, per usare un'espressione di Poe in *The Bells*, si elevano «più in alto, più in alto, sempre più in alto con un disperato desiderio.» Il centesimo numero di Bell ha 126 cifre.

Le formule per ottenere l'n-esimo numero di Bell sono complicate e difficili da usare per calcolare le serie, ma fortunatamente esiste una semplice procedura ricorsiva che ce li fornisce rapidamente e che si può capire meglio considerando la struttura del triangolo di numeri che si vede in alto a destra nella figura in basso a pagina 110. (Seguendo il suggerimento di Jeffrey Shallit lo chiamerò il triangolo di Bell.) Si comincia con un 1 al vertice e

Modi per disporre due oggetti su due** (a sinistra) **e su tre** (a destra) **piatti considerati distinti.

Il terzo numero di Bell ($B_3 = 5$) indica i modi di disporre tre oggetti su tre piatti considerati identici.

B_0	1
B_1	1
B_2	2
B_3	5
B_4	15
B_5	52
B_6	203
B_7	877
B_8	4,140
B_9	21,147
B_{10}	115,975
B_{11}	678,570
B_{12}	4,213,597

1						
1	2					
2	3	5				
5	7	10	15			
15	20	27	37	52		
52	67	87	114	151	203	
203	255	322	409	523	674	877
877	...					
⋮						

1	2	5	15	52	203	877	..
1	3	10	37	151	674	...	
2	7	27	114	523	...		
5	20	87	409	...			
15	67	322	...				
52	255	...					
203	...						

I primi numeri di Bell **(a sinistra)** *e due forme del triangolo di Bell* **(a destra).**

sotto di esso mette un altro 1. Dato che 1 più 1 è uguale a 2, si mette un 2 sulla seconda riga. Si riporta poi il 2 all'inizio della terza riga. La somma tra 2 e il numero sopra è 3, così si mette 3 a destra del 2. La somma tra 3 e il numero sopra è 5, così 5 va a destra del 3. Si continua in questo modo osservando le due regole seguenti: l'ultimo numero di ogni riga è il primo numero della riga successiva e tutti i numeri sono ottenuti sommando il numero a sinistra di quello che si vuole ottenere con il numero sopra di esso. La successione dei numeri di Bell compare ai due lati del triangolo. Se si fa ruotare il triangolo nel modo che si vede in basso a destra nella figura qui in basso, si ottiene un triangolo di differenze: ogni numero appartenente alle righe successive a quella iniziale è uguale alla differenza tra i due numeri sopra di esso.

Come il più familiare triangolo di Pascal, il triangolo di Bell è ricco di interessanti proprietà. Nel triangolo di Bell che si vede a destra in alto nella figura, le somme delle righe orrizzontali sono i numeri della seconda diagonale infinita. Se la somma di una riga è addizionata al numero di Bell con cui quella riga termina, si ottiene il numero di Bell successivo. Se al posto di ogni numero si mette un *D* quando esso è dispari e un *P* quando è pari, è facile vedere che ogni terzo numero di Bell è pari. Quindi il rapporto tra Bell dispari e Bell pari è 2:1 e la somma di ogni tripla adiacente di numeri di Bell deve essere pari.

Decine di curiose proprietà della successione di Bell sono state messe in luce e altre ancora se ne stanno scoprendo. Ad esempio, la «congruenza di Touchard» stabilisceche: $B_{p+k} \equiv B_k + B_{k+1}$ (modulo p), dove p è un numero primo. In altre parole, se l'n di B_n è espresso come la somma tra un numero primo p e un numero k, e B_n viene diviso per p, il resto sarà uguale al resto ottenuto dalla divisione per p della somma di B_k e B_{k+1}. Se k è uguale a zero, la congruenza diviene $B_p = 2$ (modulo p). In altre parole, ogni B_n per il quale n è un numero primo ha un resto di 2 quando è diviso per quel numero primo. Ad esempio, B_{13} è uguale a 27 644 437; dividendo questo numero per 13 il resto è 2.

I numeri di Bell svolgono un ruolo importante nella teoria dei numeri primi perché indicano il numero di modi in cui può essere scomposto in fattori ogni numero con fattori primi distinti. Ad esempio, 30 ha tre differenti fattori primi: 2, 3 e 5. B_3 è uguale a 5. I cinque modi per scomporre in fattori 30 sono 2 × 3 × 5, 5 × 6, 3 × 10, 2 × 15 e 30. Non è difficile vedere come questo problema sia isomorfo a quello consistente nel disporre tre oggetti distinti su tre piatti considerati identici. Si noti che tre dei primi 10 numeri di Bell sono primi: 2, 5 e 877. Ci sono altri numeri di Bell primi? Esiste un massimo tra i numeri di Bell primi? Sono due interrogativi a cui non so rispondere.

Un'altra sorprendente applicazione dei numeri di Bell: essi forniscono il numero di possibili schemi di rima per una strofa di poesia. Una strofa di un verso ha un unico schema di rima, una strofa di due versi ha due schemi di rima (i versi possono rimare oppure no), una strofa di tre versi ne ha cinque, e così via. Si dice che questa proprietà venne osservata per la prima volta dal matematico inglese J. J. Sylvester, ma io non ne ho trovato cenno nel suo breve libro *The Laws of Verse*. Sarò grato a chi mi fornirà precisazioni in proposito.

H. W. Becker, parlando dei numeri di Bell in un lavoro del 1941, introdusse quello che egli chiamò un «interessante teorema». Si chiami completamente rimata una strofa in cui ogni verso rima con almeno un altro e incompletamente rimata una strofa in cui almeno un verso non rima con nessun altro. Il numero di possibili strofe completamente rimate di n versi risulta sempre uguale al numero di possibili strofe incompletamente rimate di n-1 versi.

Henry W. Gould, un esperto di teoria dei numeri della West Virginia University, scoprì che i giapponesi avevano un interessante modo per diagrammare gli schemi di rima risalente almeno all'anno 1000 d.C. La figura qui sotto mostra i 52 diagrammi per strofe di cinque versi. Le linee verticali stanno per i versi della strofa e le linee orizzontali uniscono i versi che rimano. Gould descrisse per la prima volta i diagrammi nel 1976, nel suo *Research Bibliography of Two Special Number Sequences*, in cui vengono elencati 175 riferimenti bibliografici per i numeri di Bell e 445 per i numeri catalani. I lettori interessati possono avere una copia del libro inviando $ 3 a Gould, 1239 College Avenue, Morgantown, W. Va. 26505.

I 52 diagrammi di Murasaki, da **Storia di Genji, il principe Splendente.**

Il primo esempio di questo metodo per diagrammare gli schemi di rima si trova in *Storia di Genji, il principe splendente* (Einaudi, Torino, 1969), un famoso racconto della scrittrice giapponese Murasaki, vissuta tra il 978 e il 1031 d.C. circa. Ciascuno dei 54 capitoli del libro, tranne il primo e l'ultimo, porta come intestazione uno dei 52 diagrammi per strofe di cinque versi. Le linee verticali sono bastoncini d'incenso, ciascuno scelto tra cinque differenti colori. Le linee orizzontali uniscono bastoncini dello stesso colore. I diagrammi colorati compaiono nelle prime edizioni di questo classico giapponese, ma non nelle traduzioni inglesi. Come osservò Joanne Growney nella sua tesi di dottorato nel 1970, se si escludono i diagrammi con linee che si intersecano, il numero dei restanti diagrammi è il quinto numero catalano, 42, e questo vale in generale per diagrammi di Murasaki di n linee. Il motivo per cui Murasaki scelse questo ordine per i suoi diagrammi, scrive Gould, è sconosciuto, come la base, se ne esiste una, per ordinare i 64 esagrammi nel *I Ching*.

Ci sono numerose applicazioni dei numeri di Bell alla teoria dei grafi. Si consideri il seguente problema. Si segnino sei punti su una circonferenza, come a marcare i vertici di un invisibile esagono, e li si indichi con le lettere da a a f. Si consideri un punto isolato come un poligono convesso degenere a un vertice e due punti uniti da un segmento come un poligono convesso degenere a due vertici. Con una matita si uniscano i punti in modo da formare poligoni convessi disgiunti a uno, due, tre, quattro, cinque o sei vertici. (Disgiunti significa che due poligoni qualsiasi non possono avere un vertice in comune.) Le linee di un poligono non possono intersecarsi, perché se così fosse il poligono non sarebbe convesso; le linee di poligoni diversi, invece, possono intersecarsi. Se si vuole, si può anche non tracciare alcuna linea: si avranno così sei poligoni a un vertice. Oppure si possono unire tutti e sei i punti in modo da ottenere un unico esagono. Oppure ancora si può ottenere un miscuglio qualsiasi di poligoni; basta che siano convessi e disgiunti.

Nella figura qui sotto si possono vedere quattro possibili schemi. Quanti schemi differenti esistono? Per chi ha seguito la discussione sui numeri di Bell, il problema non presenta difficoltà. Il mese prossimo fornirò una risposta di tipo generale che si applica a n punti uniti da poligoni convessi disgiunti con un numero di vertici variante da uno a n.

Un problema di Bell.